Ferreira, M. A., Vonk, J. M., Baurecht, H., Marenholz, I., Tian, C., Hoffman, J. D., Helmer, Q., Tillander, A., Ullemar, V., van Dongen, J., Lu, Y., Rüschendorf, F., Esparza-Gordillo, J., Medway, C. W., Mountjoy, E., Burrows, K., Hummel, O., Grosche, S., Brumpton, B. M., ... the 23 and Me Research Team (2017). Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nature Genetics*, *49*(12), 1752-1757. https://doi.org/10.1038/ng.3985

Peer reviewed version

Link to published version (if available):
10.1038/ng.3985

Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology

MAIN TEXT

Manuel A Ferreira1*, Judith M Vonk2*, Hansjörg Baurecht3*, Ingo Marenholz4,5*, Chao Tian6*, Joshua D Hoffman7*, Quinta Helmer8*, Annika Tillander9*, Vilhelmina Ullemar9*, Jenny van Dongen8*, Yi Lu9*, Franz Rüschendorf 4*, Jorge Esparza-Gordillo4,5,10, Chris W Medway11, Edward Mountjoy11, Kimberley Burrows11, Oliver Hummel4, Sarah Grosche4,5, Ben M Brumpton11,12,13, John S Witte14, Jouke-Jan Hottenga8, Gonneke Willemsen8, Jie Zheng11, Elke Rodríguez3, Melanie Hotze3, Andre Franke15, Joana A Revez1, Jonathan Beesley1, Melanie C Matheson16, Shyamali C Dharmage16, Lisa M Bain1, Lars G Fritsche12, Maiken E Gabrielsen12, Brunilda Balliu17, the 23andMe Research Team6,18, AAGC collaborators19, BIOS collaborators20, LifeLines Cohort Study21, Jonas B Nielsen22,23, Wei Zhou23, Kristian Hveem12, Arnulf Langhammer24, Oddgeir L Holmen12, Mari Løset12,25, Gonçalo R Abecasis26,12, Cristen J Willer26,12,23,22, Andreas Arnold27, Georg Homuth28, Carsten O Schmidt29, Philip J Thompson30, Nicholas G Martin1, David L Duffy1, Natalija Novak31, Holger Schulz32,33, Stefan Karrasch32,34, Christian Gieger35, Konstantin Strauch36, Ronald B Melles37, David A Hinds6, Norbert Hübner4§, Stephan Weidinger3§, Patrik KE Magnusson9§, Rick Jansen38§, Eric Jorgenson37§, Young-Ae Lee4,5§, Dorret I Boomsma8§, Catarina Almqvist 9,39§, Robert Karlsson9§, Gerard H Koppelman40§ and Lavinia Paternoster11§

* These authors contributed equally to this work.

§ These authors jointly supervised this work.

Affiliations

1 Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Australia

2 Epidemiology, University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD, Groningen, the Netherlands

3 Department of Dermatology, Allergology and Venereology, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany

4 Max Delbrück Center (MDC) for Molecular Medicine, Berlin, Germany

5 Clinic for Pediatric Allergy, Experimental and Clinical Research Center of Charité Universitätsmedizin Berlin and Max Delbrück Center, Berlin, Germany

6 Research, 23andMe, Mountain View, California, USA

7 Department of Epidemiology and Biostatistics, University of Calfornia San Francisco, San Francisco, California, USA

8 Department Biological Psychology, Netherlands Twin Register , Vrije University, Amsterdam, The Netherlands

9 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

10 Current address: GlaxoSmithKline, Stevenage, UK

11 MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK

12 K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

13 Department of Thoracic Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

14 Epidemiology and Biostatistics, University of Calfornia San Francisco, San Francisco, California, USA

15 Institute of Clinical Molecular Biology, Christian Albrechts University of Kiel, Kiel, Germany

16 Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

17 Department of Pathology, Stanford University School of Medicine, Stanford, USA

18 the 23andMe Research Team is listed in Supplementary Information section 2

19 AAGC collaborators are listed in Supplementary Information section 3

20 BIOS Consortium is listed in Supplementary Information section 4

21 LifeLines Cohort Study is listed in Supplementary Information section 5

22 Department of Human Genetics, University of Michigan, Ann Arbor, USA

23 Department of Internal Medicine, University of Michigan, Ann Arbor, USA

24 The HUNT Research Centre, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

25 Department of Dermatology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

26 Center for Statistical Genetics, University of Michigan, Ann Arbor, USA

27 Clinic and Polyclinic of Dermatology, University Medicine Greifswald, Greifswald, Germany

28 Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine and Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany

29 Institute for Community Medicine, Study of Health in Pomerania/KEF, University Medicine Greifswald, Greifswald, Germany

30 Institute for Respiratory Health, Harry Perkins Institute of Medical Research, University of Western Australia, Nedlands, Australia

31 Department of Dermatology and Allergology, University-Hospital Bonn, Bonn, Germany

32 Institute of Epidemiology I, Helmholtz Zentrum Munchen - German Research Center for Environmental Health , Neuherberg, Germany

33 Comprehensive Pneumology Center Munich (CPC-M), Member of the German Center for Lung Research, Munich, Germany

34 Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Ludwig-Maximilians-Universität, Munich, Germany

35 Research Unit of Molecular Epidemiology and Institute of Epidemiology II,, Helmholtz Zentrum Munchen - German Research Center for Environmental Health , Neuherberg, Germany

36 Institute of Genetic Epidemiology , Helmholtz Zentrum Munchen - German Research Center for Environmental Health , Neuherberg, Germany

37 Division of Research, Kaiser Permanente Northern California, Oakland, California, USA

38 Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands

39 Pediatric Allergy and Pulmonology Unit at Astrid Lindgren Children’s Hospital, Karolinska University Hospital, Stockholm, Sweden

40 Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD, Groningen, the Netherlands

Corresponding author:

Manuel A R Ferreira, PhD

QIMR Berghofer Medical Research Institute

Locked Bag 2000, Royal Brisbane Hospital,

Herston QLD 4029, Australia

Phone: +61 7 3845 3552

Fax: +61 7 3362 0101

Email: manuel.ferreira@qimrberghofer.edu.au

Title character count: 89

Introductory paragraph word count: 150

Main text word count: 2,014

Display items: 3 figures and 3 tables

Reference count: 19

Online Methods word count: 4,804 (in separate Word file)

Supplementary Information: Six Supplementary Figures and Supplementary Text (in separate Word file)

Supplementary Tables: 28 tables (in separate Excel file)

INTRODUCTORY PARAGRAPH

Asthma, hay fever (or allergic rhinitis) and eczema (or atopic dermatitis) often coexist in the same individuals[1], partly because of a shared genetic origin[2-4]. To identify shared risk variants, we performed a genome-wide association study (GWAS, n=360,838) of a broad allergic disease phenotype that considers the presence of any one of these three diseases. We identified 136 independent risk variants ($P<3x10^{-8}$), including 88 not previously reported, which implicate 132 nearby genes in allergic disease pathophysiology. Disease-specific effects were detected for only six variants, confirming that most represent shared risk factors. Tissue-specific heritability and biological process enrichment analyses suggest that shared risk variants influence lymphocyte-mediated immunity. Six target genes provide an opportunity for drug repositioning, while for 36 genes CpG methylation was found to influence transcription independently of genetic effects. Asthma, hay fever and eczema partly coexist because they share many genetic risk variants that dysregulate the expression of immune-related genes.

MAIN TEXT

The analytical approach used is summarized in **Supplementary Fig. 1**. We tested for association with allergic disease 8,307,659 genetic variants that passed quality control filters (**Supplementary Table 1)**, comparing 180,129 cases who reported having suffered from asthma and/or hay fever and/or eczema, and 180,709 controls who reported not suffering from any of these diseases (**Supplementary Table 2**), all of European ancestry. Meta-analysis of results from the 13 contributing studies (**Supplementary Fig. 2**) identified 99 genomic regions (*i.e.* loci) located >1 Mb apart containing at least one genetic variant associated with allergic disease at a genome-wide significance threshold of $3x10^{-8}$ (**Fig. 1** and **Supplementary Table 3).** Based on approximate conditional analysis[5], 136 genetic variants in these 99 loci had a statistically independent association with disease risk (**Table 1**). Henceforth, we refer to these as "sentinel risk variants", which either represent, or are in linkage disequilibrium (LD) with, a causal functional variant. These included 69 (in 35 loci) located <1 Mb from risk variants reported in previous GWAS of allergic disease (**Supplementary Table 4**). Of note, 21/69 sentinel variants were in low linkage disequilibrium (LD, r^2<0.05) with the previously reported risk variants, indicating that they represent novel associations in these loci. The remaining 67 sentinel variants (in 64 loci) were located >1Mb from previously reported associations (**Supplementary Table 5**), of which 23 were in low LD with nearby variants reported for other diseases or traits (**Supplementary Table 6**). Eighteen loci had multiple independent association signals (**Supplementary Table 3**). Altogether, we identified 88 (67+21) genetic associations with allergic disease that are new, a substantial increment over the 75 associations reported previously (**Supplementary Fig. 3 and Supplementary Table 7**).

As expected from a study design that maximized power to identify shared risk variants[6], we found that 130 of the 136 sentinel variants had similar allele frequencies in case-only association analyses that

compared three non-overlapping groups of adults: those who reported suffering from asthma only (*n*=12,268), hay fever only (*n*=33,305) or eczema only (*n*=6,276) (**Supplementary Table 8**). There was thus no evidence that these 130 variants have differential effects on the three individual diseases. The six variants with evidence for stronger effects in one allergic disease when compared to the other two were located in five known allergy risk loci (*e.g. FLG* and *GSDMB*, **Fig. 2**). On the other hand, many sentinel variants (26 or 19%) were also associated with the age at which symptoms of any allergic disease first developed (*n*=35,972, **Supplementary Table 9**), the allele associated with a higher disease risk being always associated with earlier age-of-onset (**Supplementary Fig. 4**). For 18 of those 26 variants, the effect on age-of-onset was not significantly different between individual diseases (**Supplementary Table 9**), suggesting that they influence the age at which symptoms first develop for all three diseases.

We then used LD-score regression analysis[7] (see Methods) to quantify the liability-scale heritability of the three individual diseases that was collectively explained by the 136 top associations in the Nord-Trøndelag Health Study (HUNT, up to *n*=20,350), which was not part of the discovery meta-analysis. This was found to be 3.2% for asthma, 3.8% for hay fever and 1.2% for eczema, respectively representing about a fifth, a sixth and a tenth of the overall heritability of each disease that is explained by common single nucleotide polymorphisms (SNPs; **Supplementary Table 10**). Therefore, the inheritance of risk alleles at these loci partly explains why these three conditions coexist.

To understand the biological consequences of allergy risk variants, we then identified plausible target genes of the 136 sentinel variants. There were 5,739 transcripts annotated near (+/- 1 Mb) sentinel variants, including 2,569 protein-coding genes. For 132 of these transcripts, the nearby sentinel variant

was in high LD ($r^2 \geq 0.8$) with either a non-synonymous SNP (22 genes; **Supplementary Table 11**) or a sentinel expression quantitative trait locus (eQTL) identified in relevant tissues or cell types (additional 110 genes; **Supplementary Tables 12** and **13)**. We refer to these 132 transcripts as plausible target genes, which were located in 54 of the 99 risk loci (**Fig. 1** and **Supplementary Table 14**). Studies that confirm the target gene predictions and identify the underlying functional variants are warranted; genes that could be prioritized for functional follow-up include 78 identified using a more conservative LD threshold ($r^2 \geq 0.95$; **Supplementary Table 14**) or 61 predicted to be the likely targets based on independent evidence from publicly available functional data (**Supplementary Tables 15** and **16**; see Methods for details**)**. Of note, 77 (58%) of the 132 plausible target genes have not previously been implicated in allergic disease (**Supplementary Tables 14)**, and so potentially represent novel key contributors to disease pathophysiology (examples in **Table 2**).

Next, based on data from the GTEx consortium[8], we identified broad tissue types in which the plausible target genes were disproportionally expressed, using the Tissue Specific Expression Analysis (TSEA) approach described previously[9]. We excluded genes located in the major histocompatibility complex (MHC) or not present in the TSEA GTEx database, leaving 112 plausible target genes for analysis. When compared to the remaining 17,671 non-MHC genes in the genome, we found that the list of plausible targets was enriched for genes specifically expressed in whole-blood and lung (**Fig. 2A**). Both associations remained significant (**Supplementary Fig. 5**) after restricting the background gene list to the subset of 12,804 non-MHC genes with eQTLs reported in the same studies used to identify the plausible target genes (**Supplementary Tables 12)**. These results indicate that the plausible targets are enriched for genes preferentially expressed in whole-blood and lung, and that this is unlikely to arise because the plausible targets were also enriched for genes with eQTLs in those tissues.

The enrichment in whole-blood and lung expression could be a general feature of arbitrary genes located near the sentinel risk variants. To address this possibility, we determined how often the enrichment observed with the plausible target genes was exceeded when analyzing 1,000 lists of random genes. When genes were randomly selected from the same 98 non-MHC allergy risk loci identified in the meta-analysis, matching on the number of plausible target genes identified per locus (range 0 to 11) and in total (*i.e.* 112), the enrichment observed in whole-blood was not exceeded in any of the 1,000 random lists when considering results for all 25 tissues tested (**Fig. 3A** and **Supplementary Table 17**). Similar results were observed for lung. For comparison, arbitrary genes were also selected from 2 Mb loci drawn at random from the genome, or simply from all genes in the genome, and results were very similar (**Fig. 3A** and **Supplementary Table 17)**. Randomly selecting genes from the subset with eQTLs also had no impact on the results (**Supplementary Fig. 5)**. Therefore, we conclude that the enrichment in expression observed in whole-blood and lung was specific to the genes identified as plausible targets of sentinel risk variants.

To identify specific cell types that were likely to contribute to the enrichment in whole-blood, we used an orthogonal approach[10] that quantifies tissue-specific enrichments in SNP heritability rather than in gene expression. Specifically, this approach quantifies the trait heritability that is explained by SNPs that overlap cell type-specific regulatory annotations measured by the ENCODE project in 100 different cell types. In this analysis, the strongest enrichment in SNP heritability was observed for regulatory annotations measured in helper T cells (including Th17, Th1 and Th2), regulatory T cells, $CD4^+$ and $CD8^+$ memory T cells, $CD56^+$ NK cells and $CD19^+$ B cells (**Fig. 3B and Supplementary Table 18**). These results are consistent with previous findings[11] and the widely documented

contribution of these T cell subsets to allergic responses. Similar results were obtained after removing the 136 top associations from our GWAS results (**Supplementary Fig. 6 and Supplementary Table 18)**, indicating that the observed enrichments extend beyond genome-wide significant SNPs. These results demonstrate that genetic risk variants shared between asthma, hay fever and eczema, including but not limited to the ones that reached genome-wide significance, operate to a large extent by modulating gene expression in cells of the immune system.

To help understand how the sentinel variants might influence immune cell function, we then identified biological processes over-represented amongst the plausible target genes when compared to the rest of the genes in the genome (MHC excluded), using GeneNetwork[12]. As for the analysis of tissue-specific enrichment in gene expression, for each specific biological process, we compared the enrichment observed with the list of plausible target genes with that observed with 1,000 gene lists randomly drawn from the same allergy risk loci. After correcting for the 3,770 biological processes tested, we found 35 pathways for which the enrichment observed with the plausible target genes was exceeded in <5% of the random gene lists (**Fig. 3C** and **Supplementary Table 19**). These included biological processes related to T and B cell activation, B cell proliferation and isotype switching, interleukin (IL-) 2 and IL-4 production, confirming a key role for the sentinel variants and the likely target genes on lymphocyte-mediated immunity. Other noteworthy enrichments were observed for pathways related to induction of cell death, lipid phosphorylation and NK cell differentiation.

Consistent with a widespread effect of allergy risk variants on immune cell function, many sentinel risk variants have been reported to associate with other immune-related traits, notably blood cell counts (**Supplementary Table 20**) and auto-immune diseases (**Supplementary Table 21**). The genetic

234 overlap with auto-immune diseases was not restricted to sentinel variants, as evidenced by significant
235 positive genetic correlations with celiac disease, Crohn's disease and inflammatory bowel disease
236 obtained after excluding the 136 top associations from our GWAS results (**Supplementary Table 22**).
237 Other significant genetic correlations were observed for obesity- and depression-related traits, both
238 previously suggested by twin studies[13]. The former provides support for a role of allergy risk variants in
239 the regulation of metabolic homeostasis.
240
241 We then investigated whether any of the plausible target genes identified could potentially represent a
242 new opportunity for drug repositioning, as shown by others[14]. We found that 29 genes have been or are
243 being considered as drug targets, including nine for the treatment of allergic diseases (**Supplementary**
244 **Table 23)**, four for auto-immune diseases (**Supplementary Table 24**) and 16 for other diseases
245 (**Supplementary Table 25**), mostly cancer. Therefore, for 20 genes, drugs currently in development for
246 other indications might influence biological mechanisms underlying allergic disease. For six of these
247 genes, the effect on gene expression of the allergy protective allele (**Supplementary Table 26**) and the
248 existing drug matched (**Table 3)**, suggesting that the latter might attenuate (and not exacerbate) allergy
249 symptoms, and so could be prioritized for pre-clinical testing.
250
251 Finally, based on data from the BIOS consortium[15] (*n*=2,101), we found that a substantial fraction of
252 target genes (36 or 27%) had a nearby CpG site for which methylation levels were significantly
253 correlated with mRNA levels in blood, independently of SNP effects (**Supplementary Table 27**). This
254 observation raises the possibility that environmental effects on the methylation state of these CpGs
255 might influence target gene expression and, by extension, allergic disease risk. Well powered studies
256 that address this possibility are warranted. In exploratory analyses, we tested the association between

five established risk factors for allergic disease (see Methods) and the methylation state of expression-associated CpGs for those 36 genes (n=1,211). We observed only one significant association, between smoking and the methylation state of *PITPNM2* (**Supplementary Table 28**), which was reported in a previous study[16]. These results indicate that smoking might influence the risk of allergic disease partly by modulating the methylation state of expression-associated CpGs for *PITPNM2*, a PYK2-binding protein[17] potentially involved in granulocyte function[18,19].

In conclusion, we doubled the number of risk variants for allergic disease through a large GWAS of a multi-disease phenotype defined based on information from three genetically correlated diseases, asthma, hay fever and eczema. With a few exceptions, the variants identified had similar effects on the individual disease entities. The risk variants, and their likely target genes, are predicted to influence overwhelmingly the function of immune cells. Novel drugs for allergy are proposed based on genomics-guided drug repositioning. Finally, our results raise the possibility that environmental factors such as smoking might influence allergic disease risk through modulation of target gene methylation.

ONLINE METHODS

Methods for this paper are provided in a separate file.

FIGURES

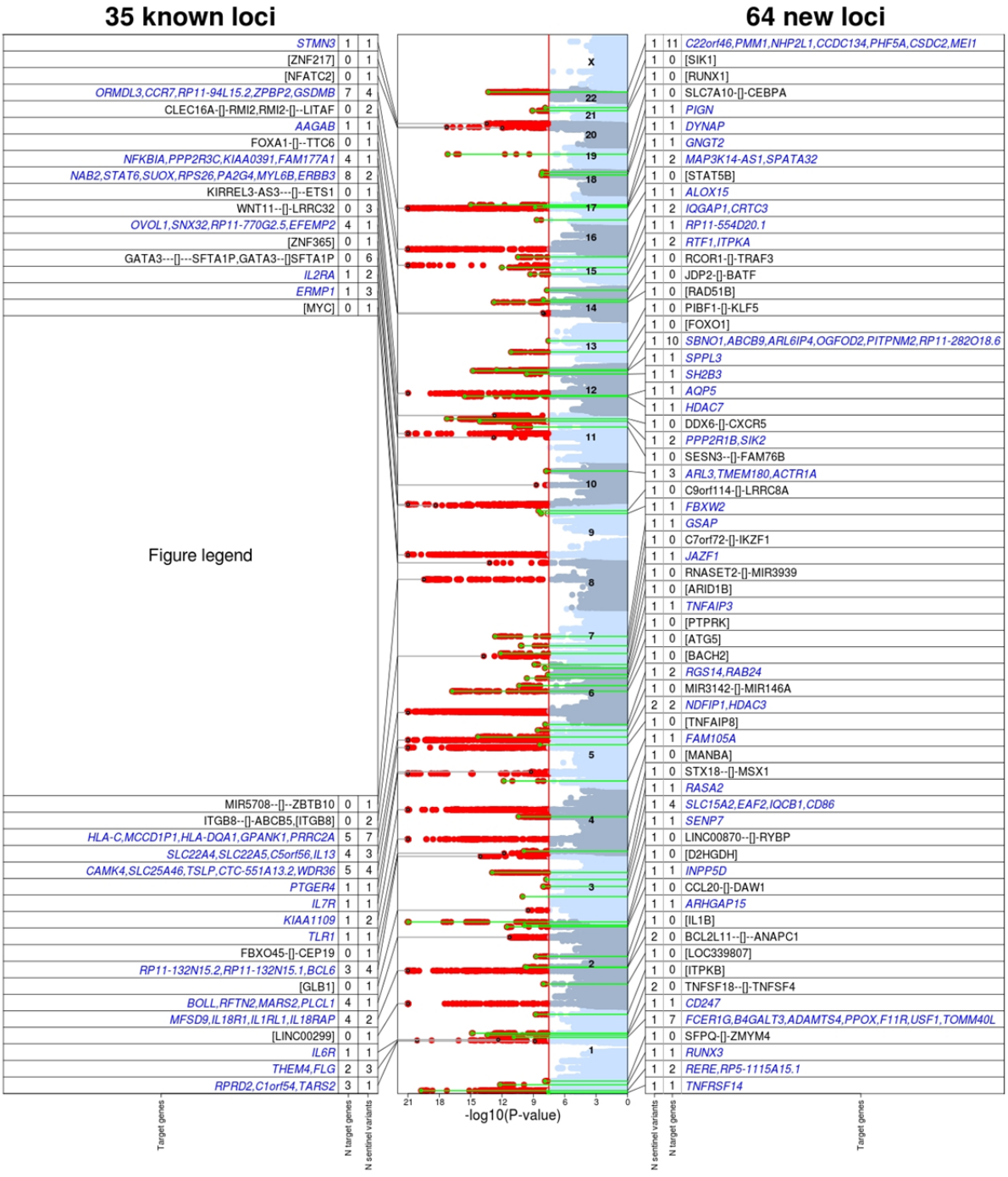


Figure 1. GWAS meta-analysis identifies 99 loci containing 136 genetic risk variants independently associated with the risk of allergic disease at $P<3x10^{-8}$. The 136 sentinel risk variants were located in 35 previously reported (69 variants) and 64 novel (67 variants) risk loci. The numbers

of plausible target genes of sentinel risk variants identified for each locus are shown, with target gene names listed in blue font. For loci with many target genes, only a selection is listed. When no target gene was identified, the nearest gene(s) to the sentinel variants are instead shown in black font. Red vertical line in Manhattan plot shows genome-wide significance threshold used (P=3x10^{-8}).

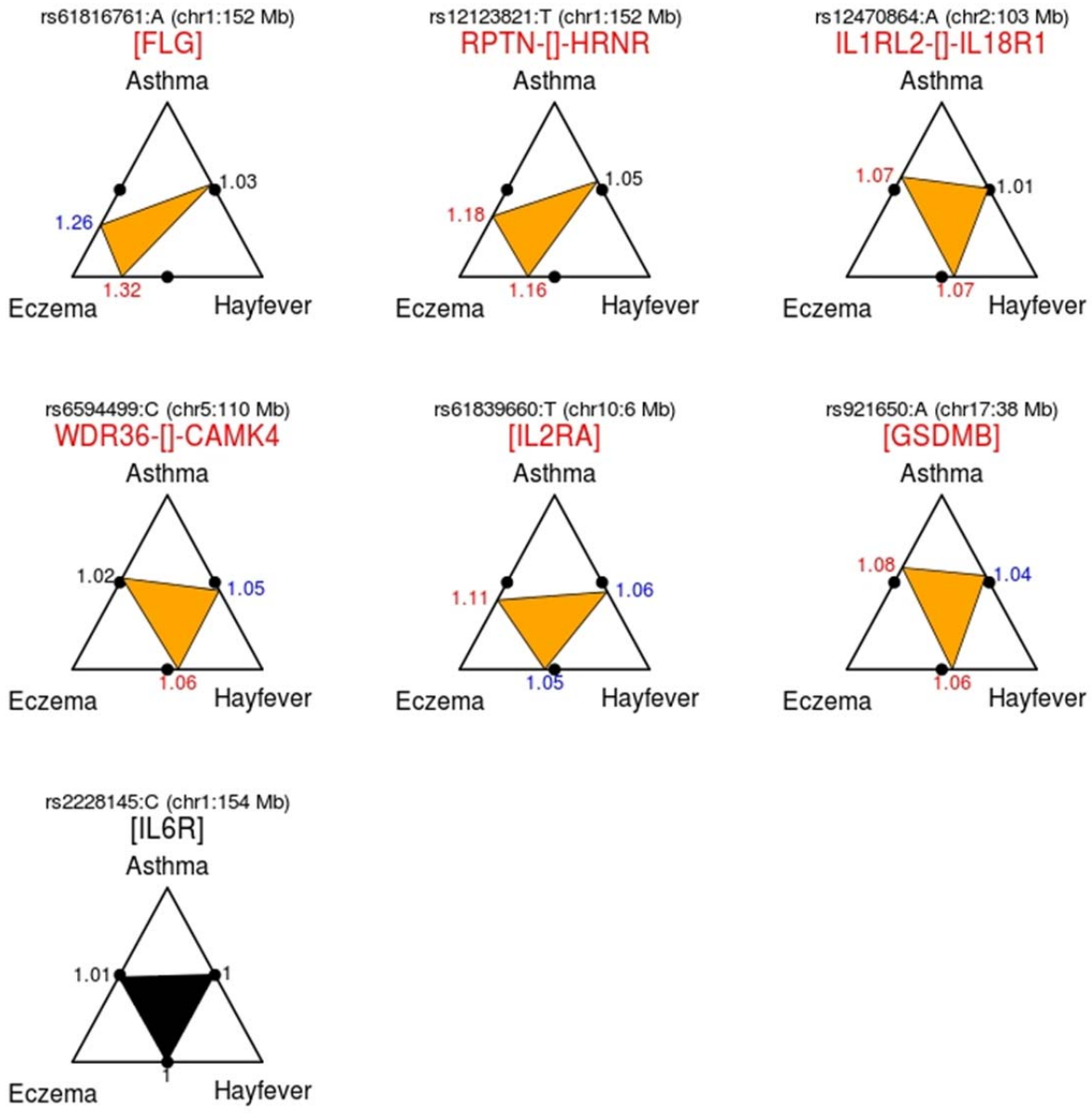
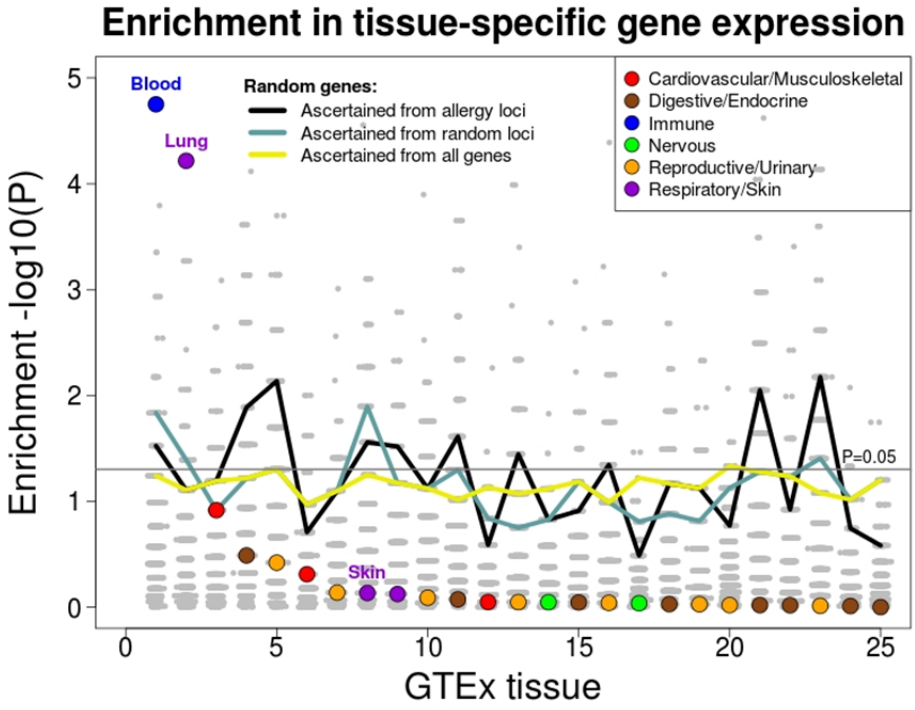


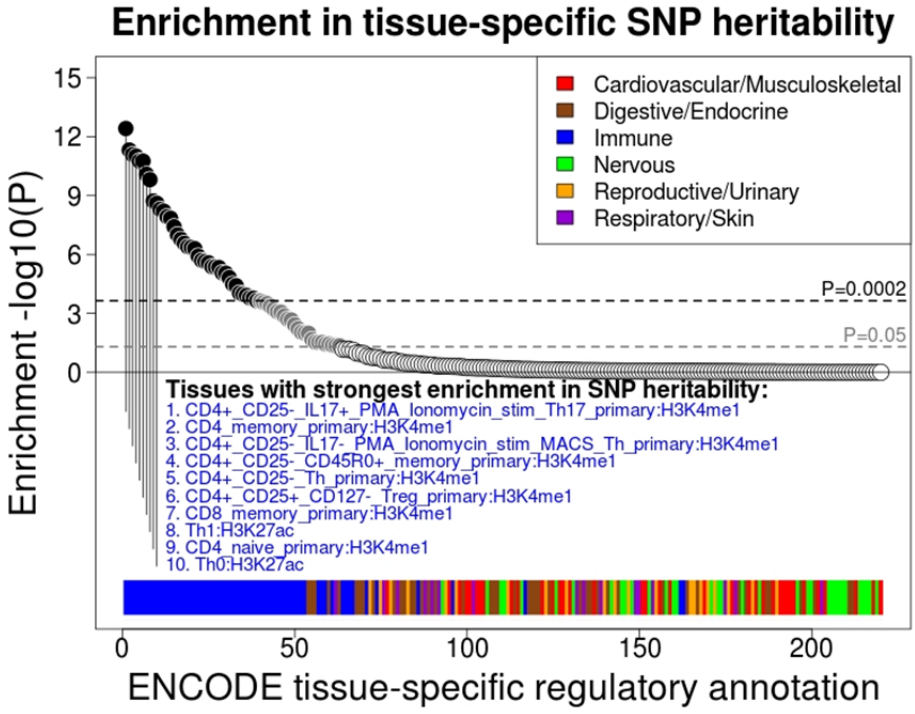
Figure 2. Six of the 136 sentinel variants had significant allele-frequency differences in pairwise case-only association analyses contrasting individuals suffering from a single allergic disease. For each sentinel variant, we performed three case-only association analyses, comparing asthma-only cases (*n*=12,268) against hay fever-only cases (*n*=33,305); asthma-only cases against eczema-only cases (*n*=6,276); and hay fever-only cases against eczema-only cases. After accounting for multiple testing, significant associations for at least one of these analyses were only observed for six of the 136 sentinel

290 variants, which are shown in the first two rows of the figure. For a given variant, the vertices of the
291 inner triangle point to the position along the edges of the outer triangle that corresponds to the allele
292 frequency difference observed between pairs of single-disease cases. For example, the rs61816761:A
293 allele, which is located in the Fillagrin gene, was 1.32-fold more common in individuals suffering only
294 from eczema when compared to individuals suffering only from hay fever (P=7.2x10^{-8}), consistent with
295 this SNP being a stronger risk factor for eczema than for hay fever. A similar result (OR = 1.26,
296 P=0.0004) was observed for this variant when contrasting eczema-only cases against asthma-only
297 cases. For comparison, a variant with no allele frequency differences in all three pairwise single-disease
298 association analyses is also shown (rs2228145, in the *IL6R* gene). In this case, the three estimated odds
299 ratios were approximately equal to 1. The color of the OR font reflects the significance of the
300 association: red for P<1.2x10^{-4} (correction for multiple testing), blue for P<0.05 and black for P>0.05.
301
302
303

(A)


Enrichment in tissue-specific gene expression
Blood
Lung
Skin
Random genes:
Ascertained from allergy loci
Ascertained from random loci
Ascertained from all genes
Cardiovascular/Musculoskeletal
Digestive/Endocrine
Immune
Nervous
Reproductive/Urinary
Respiratory/Skin
P=0.05
Enrichment -log10(P)
GTEx tissue

(B)


Enrichment in tissue-specific SNP heritability
Cardiovascular/Musculoskeletal
Digestive/Endocrine
Immune
Nervous
Reproductive/Urinary
Respiratory/Skin
P=0.0002
P=0.05
Tissues with strongest enrichment in SNP heritability:
1. CD4+_CD25-_IL17+_PMA_Ionomycin_stim_Th17_primary:H3K4me1
2. CD4_memory_primary:H3K4me1
3. CD4+_CD25-_IL17-_PMA_Ionomycin_stim_MACS_Th_primary:H3K4me1
4. CD4+_CD25-_CD45R0+_memory_primary:H3K4me1
5. CD4+_CD25-_Th_primary:H3K4me1
6. CD4+_CD25+_CD127-_Treg_primary:H3K4me1
7. CD8_memory_primary:H3K4me1
8. Th1:H3K27ac
9. CD4_naive_primary:H3K4me1
10. Th0:H3K27ac
Enrichment -log10(P)
ENCODE tissue-specific regulatory annotation

(C)

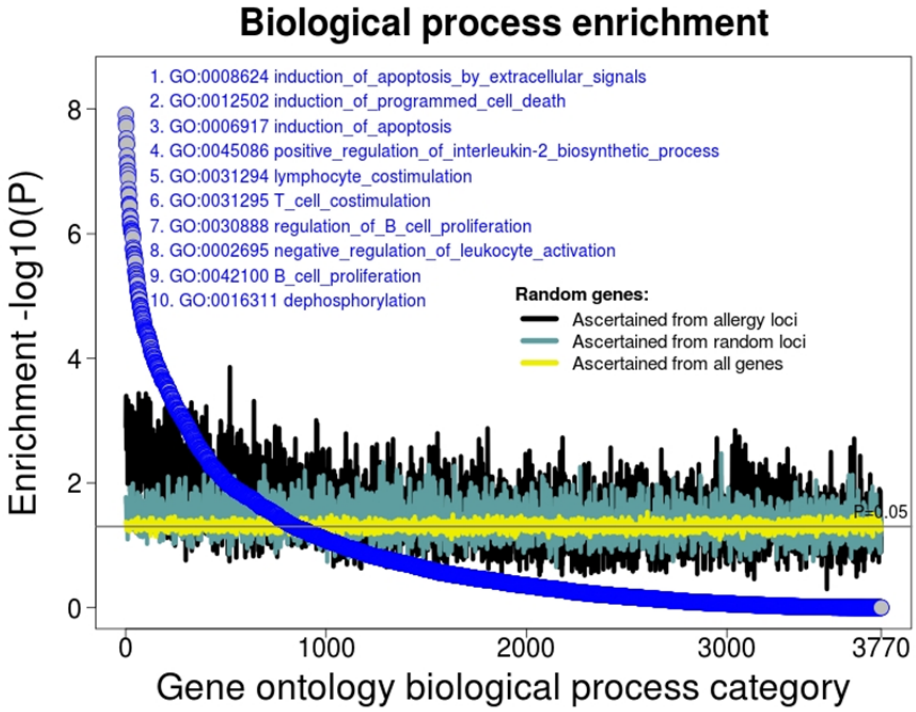


Figure 3. Tissues and biological processes influenced by allergy risk variants.

(A) Enrichment of tissue-specific gene expression in 25 broad tissues studied by the GTEx consortium. We used the TSEA approach[9] to test if genes specifically expressed in a given tissue were enriched amongst the list of plausible target genes when compared to other genes in the genome. The enrichment (*y*-axis) is shown as the -log10 of the Fisher's exact test *P*-value. For comparison, we analyzed 1,000 lists of random genes instead of the plausible target genes. We selected genes at random using three strategies (see Methods for details). First, genes were randomly drawn from the 98 non-MHC allergy risk loci identified in our GWAS, matching on the number selected per locus and in total. The enrichment *P*-value for each of the 1,000 random gene lists is shown by a grey circle. The black-solid line shows the *P*-value for the 50th most significant random list (*i.e.* corresponding to the 5th

percentile): under the null hypothesis of no enrichment, this *P*-value should be close to 0.05 (horizontal grey line). Second, genes were drawn at random from 2 Mb loci selected at random from the genome, matching on the number of genes selected (and available for selection) per locus and in total. Third, genes were drawn at random from all 18,300 genes available for analysis. For the latter two strategies, the *P*-value for the 50th most significant random gene list is shown by the blue and yellow lines, respectively; enrichment results for each individual random dataset are not shown. Similar results were obtained after restricting the random genes and the background gene list to the subset of genes with eQTLs (**Supplementary Fig. 6**). Genes in the MHC were excluded from these analyses.

.

(B) Enrichment of SNP-based heritability in 220 individual cell type-specific regulatory annotations. We used stratified LD score regression analysis [10] to quantify the contribution of SNPs that overlap cell type-specific regulatory annotations to the SNP-based disease heritability. Annotations with an enrichment in SNP heritability (-log10 of the *P*-value of the regression coefficient, *y*-axis) that was significant after correcting for multiple testing ($P<0.0002$) are shown in black circles (top 10 listed in blue font; all results in **Supplementary Table 18**). SNPs in the MHC were excluded from these analyses.

(C) Biological processes enriched amongst the list of plausible target genes. We used GeneNetwork[12] to test if the plausible target genes as a group were more likely to be part of a specific biological process category when compared to the rest of the genes in the genome. The enrichment (*y*-axis) is shown as the –log10 of the Wilcoxon rank-sum test *P*-value (see Methods for details). The top 10 pathways are listed in blue font. For comparison, we analyzed 1,000 random gene lists generated using the same three strategies described above. For each of these strategies, the *P*-value for the 50th

most significant random gene list is shown by the black (random genes from allergy loci), blue (random genes from random loci) and yellow (random genes selected from all available genes) lines. Similar results were obtained after restricting the random genes and the background gene list to the subset of genes with eQTLs (not shown). Genes in the MHC were excluded from these analyses.

TABLES

Table 1. Main association results for the 136 variants independently associated with the risk of allergic disease at $P<3x10^{-8}$.

Locus	Index in locus	Chr	Bp	SNP	EA	OA	OR	SE	P-value	P-value in conditional analyses[a]	Freq	Gene context	Novelty status
1	1	1	2510755	rs10910095	G	A	1.042	0.0073	2.70E-08	NA	0.84	TNFRSF14-[]-FAM213B	NewLocus
2	1	1	8482078	rs301806	T	C	1.048	0.0050	1.77E-20	NA	0.54	[RERE]	NewLocus
3	1	1	25251923	rs760805	T	A	1.038	0.0051	6.39E-13	NA	0.58	[RUNX3]	NewLocus
4	1	1	35681738	rs76167968	T	C	1.056	0.0093	1.26E-08	NA	0.93	SFPQ-[]-ZMYM4	NewLocus
5	1	1	150265704	rs7512552	C	T	1.031	0.0049	1.43E-09	NA	0.52	C1orf54-[]-MRPS21	KnownLocus-KnownVariant
6	1	1	152285861	rs61816761	A	G	1.224	0.0211	7.38E-21	NA	0.01	[FLG]	KnownLocus-KnownVariant
6	2	1	152179152	rs12123821	T	C	1.111	0.0124	6.81E-17	4.41E-018	0.05	RPTN-[]-HRNR	KnownLocus-KnownVariant
6	3	1	151796742	rs11204896	C	G	1.063	0.0085	2.36E-12	1.18E-010	0.90	[RORC]	KnownLocus-NewVariant
7	1	1	154426970	rs2228145	C	A	1.038	0.0050	4.30E-13	NA	0.35	[IL6R]	KnownLocus-KnownVariant
8	1	1	161185058	rs2070901	T	G	1.039	0.0056	1.31E-11	NA	0.26	NDUFS2-[]-FCER1G	NewLocus
9	1	1	167431352	rs2988277	C	T	1.040	0.0051	4.00E-14	NA	0.66	[CD247]	NewLocus
10	1	1	173146921	rs4090390	A	C	1.048	0.0058	1.32E-15	NA	0.22	TNFSF18--[]-TNFSF4	NewLocus
10	2	1	172700868	rs1102705	G	A	1.058	0.0088	3.07E-10	4.20E-011	0.11	FASLG-[]--TNFSF18	NewLocus
11	1	1	226914734	rs697852	A	G	1.041	0.0066	1.58E-09	NA	0.85	[ITPKB]	NewLocus
12	1	2	8442248	rs10174949	G	A	1.066	0.0054	7.34E-31	NA	0.68	[LINC00299]	KnownLocus-KnownVariant
13	1	2	64836267	rs4671601	C	T	1.039	0.0064	8.79E-09	NA	0.83	[LOC339807]	NewLocus
14	1	2	102941311	rs10865050	G	A	1.130	0.0073	6.98E-61	NA	0.85	[IL18R1]	KnownLocus-KnownVariant
14	2	2	102926362	rs12470864	A	G	1.057	0.0051	4.23E-26	3.63E-010	0.35	IL1RL2-[]-	KnownLocus-

												IL18R1	KnownVariant
15	1	2	112388538	rs4848612	A	G	1.038	0.0058	2.33E-10	NA	0.76	BCL2L11--[]--ANAPC1	NewLocus
15	2	2	112269127	rs13403656	A	T	1.045	0.0077	2.18E-08	2.84E-008	0.16	BCL2L11--[]--ANAPC1	NewLocus
16	1	2	113590467	rs1143633	C	T	1.034	0.0052	1.65E-10	NA	0.70	[IL1B]	NewLocus
17	1	2	143831599	rs74847330	A	G	1.048	0.0076	1.76E-09	NA	0.88	KYNU-[]-ARHGAP15	NewLocus
18	1	2	198950240	rs1064213	G	A	1.035	0.0049	5.37E-12	NA	0.47	[PLCL1]	KnownLocus-KnownVariant
19	1	2	228707862	rs13384448	T	C	1.042	0.0058	2.82E-12	NA	0.76	CCL20-[]-DAW1	NewLocus
20	1	2	234115629	rs1057258	C	T	1.045	0.0067	1.39E-10	NA	0.80	[INPP5D]	NewLocus
21	1	2	242698640	rs34290285	G	A	1.082	0.0064	4.05E-33	NA	0.73	[D2HGDH]	NewLocus
22	1	3	33069091	rs6776757	G	A	1.033	0.0050	3.14E-10	NA	0.47	[GLB1]	KnownLocus-KnownVariant
23	1	3	72394852	rs61192126	T	C	1.037	0.0055	8.85E-11	NA	0.69	LINC00870--[]-RYBP	NewLocus
24	1	3	101242751	rs13088318	A	G	1.031	0.0053	8.63E-09	NA	0.62	FAM172BP-[]-TRMT10C	NewLocus
25	1	3	121652141	rs75557865	G	A	1.029	0.0050	1.63E-08	NA	0.57	[SLC15A2]	NewLocus
26	1	3	141321836	rs10663129	ACT	A	1.042	0.0054	1.12E-13	NA	0.33	[RASA2]	NewLocus
27	1	3	188133336	rs60946162	T	C	1.041	0.0051	8.57E-15	NA	0.42	[LPP]	KnownLocus-KnownVariant
27	2	3	188402586	rs17607589	C	T	1.053	0.0066	1.83E-14	2.50E-015	0.84	[LPP]	KnownLocus-NewVariant
27	3	3	187633268	rs519973	A	G	1.034	0.0052	4.50E-10	2.57E-011	0.35	BCL6--[]--LPP-AS2	KnownLocus-NewVariant
27	4	3	187793833	rs2030030	T	C	1.041	0.0068	1.01E-08	9.55E-009	0.86	BCL6--[]-LPP-AS2	KnownLocus-NewVariant
28	1	3	196372546	rs80064395	C	T	1.070	0.0094	1.55E-12	NA	0.94	FBXO45-[]-CEP19	KnownLocus-NewVariant
29	1	4	4775401	rs10033073	G	A	1.040	0.0059	1.19E-10	NA	0.35	STX18--[]-MSX1	NewLocus
30	1	4	38798648	rs5743618	C	A	1.100	0.0058	3.29E-58	NA	0.70	[TLR1]	KnownLocus-KnownVariant
31	1	4	103593898	rs227275	C	A	1.034	0.0050	3.69E-11	NA	0.56	[MANBA]	NewLocus
32	1	4	123316076	rs4145717	T	G	1.059	0.0052	9.18E-27	NA	0.34	[ADAD1]	KnownLocus-

													KnownVariant
32	2	4	123454110	rs150254607	AT AT	A	1.078	0.0097	4.63E-14	2.08E-024	0.07	IL2-[]-IL21	KnownLocus-NewVariant
33	1	5	14610309	rs16903574	G	C	1.071	0.0095	1.40E-12	NA	0.08	[FAM105A]	NewLocus
34	1	5	35862841	rs7717955	C	T	1.073	0.0055	9.13E-36	NA	0.71	[IL7R]	KnownLocus-KnownVariant
35	1	5	40492655	rs7714574	T	C	1.032	0.0050	5.94E-10	NA	0.53	DAB2---[]--PTGER4	KnownLocus-KnownVariant
36	1	5	110470137	rs6594499	C	A	1.075	0.0050	4.64E-46	NA	0.48	WDR36-[]-CAMK4	KnownLocus-KnownVariant
36	2	5	110166083	rs6869502	T	A	1.079	0.0066	6.38E-29	4.46E-029	0.18	SLC25A46-[]--TSLP	KnownLocus-KnownVariant
36	3	5	110401872	rs1837253	C	T	1.070	0.0056	1.63E-31	7.28E-020	0.78	SLC25A46--[]-TSLP	KnownLocus-KnownVariant
36	4	5	110159879	rs1814576	C	T	1.121	0.0114	1.41E-22	2.05E-012	0.09	SLC25A46-[]--TSLP	KnownLocus-KnownVariant
37	1	5	118684297	rs250308	T	C	1.031	0.0051	3.95E-09	NA	0.37	[TNFAIP8]	NewLocus
38	1	5	131996500	rs848	A	C	1.068	0.0063	1.52E-24	NA	0.24	[IL13]	KnownLocus-KnownVariant
38	2	5	131799626	rs3749833	C	T	1.039	0.0056	3.30E-11	2.67E-009	0.30	[C5orf56]	KnownLocus-NewVariant
38	3	5	131989136	rs3091307	G	A	1.062	0.0062	3.64E-21	4.40E-010	0.20	RAD50-[]-IL13	KnownLocus-KnownVariant
39	1	5	141494934	rs10068717	T	C	1.042	0.0052	4.80E-15	NA	0.62	[NDFIP1]	NewLocus
39	2	5	140925362	rs740474	C	T	1.034	0.0051	5.62E-11	6.73E-011	0.42	[DIAPH1]	NewLocus
40	1	5	159909345	rs2910162	G	A	1.033	0.0053	2.55E-09	NA	0.68	MIR3142-[]-MIR146A	NewLocus
41	1	5	176782218	rs13153019	C	T	1.035	0.0059	1.33E-08	NA	0.26	LMAN2-[]-RGS14	NewLocus
42	1	6	32626403	rs34004019	A	G	1.101	0.0062	3.78E-52	NA	0.70	HLA-DQA1-[]-HLA-DQB1	KnownLocus-KnownVariant
42	2	6	31323012	rs2854001	A	G	1.059	0.0062	1.18E-19	5.46E-019	0.18	[HLA-B]	KnownLocus-KnownVariant
42	3	6	31351664	rs2507978	G	A	1.035	0.0052	1.23E-10	2.43E-015	0.41	HLA-B-[]-MICA	KnownLocus-NewVariant
42	4	6	33647058	rs10947428	C	T	1.046	0.0061	3.54E-13	6.45E-015	0.19	[ITPR3]	KnownLocus-KnownVariant
42	5	6	29893575	rs9259819	G	T	1.036	0.0058	2.44E-09	3.45E-010	0.50	[HLA-J]	KnownLocus-

													NewVariant
42	6	6	31574525	rs28895016	C	T	1.097	0.0113	9.36E-16	2.48E-009	0.91	NCR3-[]-AIF1	KnownLocus-KnownVariant
42	7	6	33046752	rs3097670	G	C	1.064	0.0089	7.73E-12	2.02E-008	0.88	[HLA-DPA1]	KnownLocus-NewVariant
43	1	6	90987512	rs2134814	C	G	1.046	0.0052	1.65E-17	NA	0.65	[BACH2]	NewLocus
44	1	6	106667535	rs9372120	G	T	1.042	0.0062	4.23E-11	NA	0.18	[ATG5]	NewLocus
45	1	6	128294709	rs35469349	A	T	1.037	0.0056	2.32E-10	NA	0.26	[PTPRK]	NewLocus
46	1	6	138195151	rs5029937	G	T	1.081	0.0136	2.37E-08	NA	0.97	[TNFAIP3]	NewLocus
47	1	6	157419508	rs9383820	C	T	1.037	0.0063	1.24E-08	NA	0.76	[ARID1B]	NewLocus
48	1	6	167390671	rs72033857	C	CCTTT	1.057	0.0090	1.25E-09	NA	0.11	RNASET2-[]-MIR3939	NewLocus
49	1	7	20560996	rs6461503	T	C	1.039	0.0049	1.75E-14	NA	0.53	ITGB8--[]-ABCB5	KnownLocus-NewVariant
49	2	7	20376018	rs10486391	A	G	1.030	0.0050	6.82E-09	1.46E-009	0.53	[ITGB8]	KnownLocus-NewVariant
50	1	7	28156887	rs6977955	T	C	1.046	0.0061	7.12E-13	NA	0.21	[JAZF1]	NewLocus
51	1	7	50253897	rs17664743	A	G	1.042	0.0061	6.22E-11	NA	0.20	C7orf72-[]-IKZF1	NewLocus
52	1	7	77018542	rs4296977	C	T	1.055	0.0071	2.14E-13	NA	0.16	[GSAP]	NewLocus
53	1	8	81292599	rs7824394	A	C	1.050	0.0052	3.47E-20	NA	0.37	MIR5708--[]--ZBTB10	KnownLocus-KnownVariant
54	1	8	128814091	rs6990534	A	G	1.042	0.0054	6.35E-14	NA	0.36	[MYC]	KnownLocus-KnownVariant
55	1	9	6208030	rs144829310	T	G	1.090	0.0068	1.19E-35	NA	0.16	RANBP6--[]-IL33	KnownLocus-KnownVariant
55	2	9	6051399	rs343478	G	A	1.033	0.0050	2.59E-10	8.68E-015	0.52	RANBP6-[]--IL33	KnownLocus-NewVariant
55	3	9	5064193	rs16922576	C	T	1.036	0.0056	3.22E-10	3.90E-011	0.32	[JAK2]	KnownLocus-NewVariant
56	1	9	123650534	rs10760123	T	G	1.032	0.0053	5.23E-09	NA	0.38	PHF19-[]-TRAF1	NewLocus
57	1	9	131613191	rs12551834	G	A	1.058	0.0093	3.02E-09	NA	0.92	C9orf114-[]-LRRC8A	NewLocus
58	1	10	6094697	rs61839660	T	C	1.080	0.0085	4.42E-19	NA	0.07	[IL2RA]	KnownLocus-KnownVariant
58	2	10	6074451	rs4747846	C	G	1.036	0.0051	1.00E-11	3.69E-009	0.52	[IL2RA]	KnownLocus-KnownVariant

59	1	10	9049253	rs12413578	C	T	1.095	0.0082	1.30E-27	NA	0.90	GATA3---[]---SFTA1P	KnownLocus-KnownVariant
59	2	10	9064361	rs1444789	C	T	1.066	0.0064	1.48E-22	5.05E-018	0.20	GATA3---[]---SFTA1P	KnownLocus-KnownVariant
59	3	10	8605553	rs11255753	T	G	1.041	0.0056	2.02E-12	6.97E-014	0.26	GATA3--[]SFTA1P	KnownLocus-NewVariant
59	4	10	9032555	rs72782676	C	G	1.303	0.0390	3.16E-11	6.29E-012	0.98	GATA3---[]---SFTA1P	KnownLocus-NewVariant
59	5	10	8841669	rs2025758	T	C	1.041	0.0050	4.68E-15	9.41E-012	0.58	GATA3---[]---SFTA1P	KnownLocus-NewVariant
59	6	10	8936162	rs11255968	C	T	1.090	0.0147	7.74E-09	2.92E-010	0.96	GATA3---[]---SFTA1P	KnownLocus-KnownVariant
60	1	10	64382359	rs2893907	C	A	1.031	0.0050	1.84E-09	NA	0.58	[ZNF365]	KnownLocus-KnownVariant
61	1	10	104225832	rs10883723	C	T	1.030	0.0052	1.58E-08	NA	0.31	C10orf95-[]-ACTR1A	NewLocus
62	1	11	65551957	rs479844	G	A	1.038	0.0050	1.60E-13	NA	0.57	AP5B1-[]-OVOL1	KnownLocus-KnownVariant
63	1	11	76293758	rs7936323	A	G	1.088	0.0049	2.24E-63	NA	0.46	WNT11--[]-LRRC32	KnownLocus-KnownVariant
63	2	11	76299431	rs55646091	A	G	1.179	0.0122	2.26E-40	4.20E-023	0.05	WNT11--[]-LRRC32	KnownLocus-KnownVariant
63	3	11	76343428	rs11236814	A	T	1.068	0.0085	3.98E-14	9.81E-012	0.91	WNT11--[]-LRRC32	KnownLocus-NewVariant
64	1	11	95425526	rs59593577	C	T	1.053	0.0075	1.58E-11	NA	0.87	SESN3--[]-FAM76B	NewLocus
65	1	11	111470567	rs7130753	C	T	1.045	0.0056	7.05E-15	NA	0.69	LAYN-[]-SIK2	NewLocus
66	1	11	118743286	rs12365699	G	A	1.061	0.0067	5.15E-18	NA	0.85	DDX6-[]-CXCR5	NewLocus
67	1	11	128158189	rs56129466	A	G	1.047	0.0061	1.92E-13	NA	0.79	KIRREL3-AS3---[]---ETS1	KnownLocus-KnownVariant
68	1	12	48196982	rs55726902	G	A	1.051	0.0060	2.59E-16	NA	0.80	[HDAC7]	NewLocus
69	1	12	50345671	rs11169225	A	T	1.045	0.0064	1.23E-11	NA	0.18	[AQP2]	NewLocus
70	1	12	57489709	rs1059513	T	C	1.084	0.0081	1.05E-22	NA	0.89	[STAT6]	KnownLocus-KnownVariant
70	2	12	56401085	rs10876864	G	A	1.047	0.0050	1.42E-19	1.99E-019	0.39	SUOX-[]-IKZF4	KnownLocus-KnownVariant

71	1	12	111932800	rs7137828	T	C	1.033	0.0050	2.22E-10	NA	0.54	[ATXN2]	NewLocus
72	1	12	121363724	rs6489785	T	C	1.043	0.0051	1.55E-15	NA	0.37	SPPL3-[]-HNF1A-AS1	NewLocus
73	1	12	123742692	rs63406760	T	TG	1.047	0.0062	2.95E-13	NA	0.77	C12orf65-[]-CDK2AP1	NewLocus
74	1	13	41173408	rs4943794	C	G	1.043	0.0061	7.18E-12	NA	0.23	[FOXO1]	NewLocus
75	1	13	73627275	rs9573092	A	G	1.030	0.0052	2.67E-08	NA	0.70	PIBF1-[]-KLF5	NewLocus
76	1	14	35761675	rs1048990	G	C	1.039	0.0066	1.04E-08	NA	0.16	[PSMA6]	KnownLocus-KnownVariant
77	1	14	38097001	rs111914382	TG	T	1.036	0.0059	8.09E-09	NA	0.27	FOXA1-[]--TTC6	KnownLocus-KnownVariant
78	1	14	68754417	rs2104047	T	C	1.042	0.0054	1.64E-13	NA	0.36	[RAD51B]	NewLocus
79	1	14	75968608	rs9323612	A	G	1.032	0.0053	8.58E-09	NA	0.70	JDP2-[]-BATF	NewLocus
80	1	14	103235012	rs9989163	A	G	1.029	0.0050	1.92E-08	NA	0.49	RCOR1-[]-TRAF3	NewLocus
81	1	15	41782684	rs12440045	C	A	1.033	0.0051	4.89E-10	NA	0.55	RTF1-[]-ITPKA	NewLocus
82	1	15	61068347	rs10519067	G	A	1.055	0.0073	9.32E-13	NA	0.86	[RORA]	NewLocus
83	1	15	67448363	rs56375023	A	G	1.073	0.0059	8.24E-32	NA	0.21	[SMAD3]	KnownLocus-KnownVariant
84	1	15	91045408	rs3540	G	A	1.036	0.0053	3.32E-11	NA	0.65	[IQGAP1]	NewLocus
85	1	16	11277358	rs11644510	C	T	1.072	0.0053	6.11E-38	NA	0.64	CLEC16A-[]-RMI2	KnownLocus-KnownVariant
85	2	16	11491007	rs12596613	C	G	1.032	0.0054	6.06E-09	2.53E-009	0.67	RMI2-[]--LITAF	KnownLocus-KnownVariant
86	1	17	4521473	rs71368508	C	A	1.124	0.0191	2.02E-09	NA	0.99	SMTNL2-[]-ALOX15	NewLocus
87	1	17	38069076	rs921650	A	G	1.059	0.0049	5.67E-30	NA	0.48	[GSDMB]	KnownLocus-KnownVariant
87	2	17	38764524	rs112401631	A	T	1.260	0.0213	2.20E-26	3.74E-026	0.01	CCR7-[]-SMARCE1	KnownLocus-NewVariant
87	3	17	38770641	rs11464691	TA	T	1.052	0.0052	1.37E-21	3.05E-020	0.57	CCR7-[]-SMARCE1	KnownLocus-NewVariant
87	4	17	38149033	rs11652139	A	G	1.051	0.0051	7.54E-22	1.70E-008	0.59	[PSMD3]	KnownLocus-KnownVariant
88	1	17	40414862	rs7207591	A	G	1.038	0.0060	1.43E-09	NA	0.76	[STAT5B]	NewLocus
89	1	17	43430696	rs7214661	G	A	1.032	0.0054	1.20E-08	NA	0.32	MAP3K14-[]-	NewLocus

												ARHGAP27	
90	1	17	47398070	rs9889262	A	T	1.043	0.0051	9.65E-16	NA	0.35	[ZNF652]	NewLocus
91	1	18	52336175	rs4801001	T	C	1.031	0.0051	5.91E-09	NA	0.42	DYNAP-[]--RAB27B	NewLocus
92	1	18	60009814	rs4574025	T	C	1.030	0.0050	6.79E-09	NA	0.56	[TNFRSF11A]	NewLocus
93	1	19	33721455	rs10414065	C	T	1.098	0.0106	6.10E-18	NA	0.92	SLC7A10-[]-CEBPA	NewLocus
94	1	20	50157837	rs3787184	A	G	1.049	0.0066	1.06E-12	NA	0.78	[NFATC2]	KnownLocus-KnownVariant
95	1	20	52208356	rs2766678	G	A	1.057	0.0063	5.04E-18	NA	0.21	[ZNF217]	KnownLocus-NewVariant
96	1	20	62322699	rs6011033	G	A	1.047	0.0060	3.50E-14	NA	0.78	[RTEL1]	KnownLocus-KnownVariant
97	1	21	36467830	rs73205303	A	G	1.044	0.0069	7.90E-10	NA	0.13	[RUNX1]	NewLocus
98	1	21	44846426	rs76081789	T	C	1.065	0.0108	1.34E-08	NA	0.94	[SIK1]	NewLocus
99	1	22	41816652	rs5758343	A	T	1.048	0.0061	4.75E-14	NA	0.22	TEF-[]-TOB2	NewLocus

EA: effect allele. OA: other allele. Freq: Effect allele frequency in Europeans populations of the 1000 Genomes Project.

[a] Eighteen loci were found to have multiple independent associations in approximate conditional analyses. We studied these 18 loci in greater detail using data from the UK Biobank (132,702 unrelated individuals of European descent) to confirm these results. Specifically, for each of these 18 loci, we tested if the independent variants identified in the GWAS meta-analysis were associated with disease risk when included simultaneously in a logistic regression model, using R (*e.g.* glm(disease ~ rs61816761 + rs12123821 + rs11204896 + covariates)). For 17 of the 18 loci, all sentinel SNPs were independently associated with disease risk at $P<0.05$ (not shown). The exception was locus #36, for which one of the 4 sentinel SNPs (rs6594499) had a $P=0.12$. Therefore, despite the reduced sample size, results from this analysis in the UK Biobank study confirm that all but one of the independent variants identified by approximate conditional analysis in these 18 loci have a statistically independent association with disease risk.

366 **Table 2. Selected examples of plausible target genes not previously implicated in the pathophysiology of allergic disease.**

Gene	Summary	Possible role(s) in allergic disease[a]
RERE	Nuclear receptor coregulator that positively regulates retinoic acid signaling	Positive regulation of B cell differentiation, eosinophil survival and migration
PPP2R3C	Sub-unit of protein phosphatase 2A (PP2A) that regulates immune cell function	Th2 differentiation, Treg function, response to viral infection
RASA2	GTPase-activating protein of Ras that regulates receptor signal transduction	Unknown. RASA3: hematopoiesis. RASA4: macrophage phagocytosis.
SIK2	Salt-inducible kinase	Regulation of macrophage inflammatory phenotype, metabolic homeostasis
RTF1	Component of the PAF complex, that is involved in transcriptional regulation	Anti-viral response, regulation of TNF expression
SMARCE1	Sub-unit of the BAF chromatin remodeling complex	Repressor of CD4 differentiation
DYNAP	Dynactin-associated protein that activates protein kinase B	Cytokine signaling, T cell function
THEM4	Mithocondrial thioesterase that is a negative regulator of protein kinase B	Vitamin D-dependent macrophage-mediated inflammation
ARHGAP15	Rho GTPase activating protein that down-regulates RAC1	Rac1-dependent inflammatory response
SENP7	Sentrin/small ubiquitin-like modifier (SUMO)-specific protease	Susceptibility to viral infection
SLC15A2	Peptide transporter expressed in skin, nasal and lung epithelial cells	Bacterial peptide recognition and immune activation

367 [a] References that support the possible role(s) listed are cited in the Supplementary Information.

Table 3. Plausible target genes with drugs in development for indications other than allergic diseases, for which the effect on gene expression of the allergy protective allele and the existing drug matched.

Plausible target gene	Effect of allergy protective allele on gene expression	Drug Action	Drug Status	Drug Name	Originator Company	Active Indications
CD86	Increased	Agonist	Discovery	BR-02001	Boryung_Pharm_Co_Ltd	Autoimmune_disease
CCR7	Decreased	Antagonist	Discovery	anti-CCR7_chimeric_IgG1_antibodies	North_Coast_Biologics_LLC	Unidentified_indication
CCR7	Decreased	Antagonist	Discovery	anti-CCR7_monoclonal_antibody	Pepscan_Systems_BV	Cancer
CCR7	Decreased	Antagonist	Discovery	CCR7-targeting_antibody	Abilita_Bio_Inc	Metastatic_breast_cancer
CCR7	Decreased	Antagonist	NA	chemokine_antagonists	Neurocrine_Biosciences_Inc	NA
CCR7	Decreased	Antagonist	NA	chemokine_receptor_inhibitors	Sosei_Group_Corp	NA
F11R	Decreased	Antagonist	Discovery	F11R_inhibitors	Provid_Pharmaceuticals_Inc	Cardiovascular_disease
F11R	Decreased	Antagonist	Discovery	F-50073	Pierre_Fabre_SA	Cancer
PHF5A	Decreased	Antagonist	Discovery	PHF5A_inhibitors	Fred_Hutchinson_Cancer_Research_Center	Glioblastoma
RGS14	Decreased	Antagonist	NA	regulator_of_G-protein_signaling_14_inhibitor	University_of_Malaga	Memory loss
TARS2	Decreased	Antagonist	Discovery	borrelidin	Scripps_Research_Institute	Infectious_disease

REFERENCES

1 Pinart, M. *et al.* Comorbidity of eczema, rhinitis, and asthma in IgE-sensitised and non-IgE-sensitised children in MeDALL: a population-based cohort study. *The Lancet. Respiratory medicine* **2**, 131-140, doi:10.1016/S2213-2600(13)70277-7 (2014).

2 Thomsen, S. F. *et al.* Findings on the atopic triad from a Danish twin registry. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* **10**, 1268-1272 (2006).

3 van Beijsterveldt, C. E. & Boomsma, D. I. Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. *Eur Respir J* **29**, 516-521, doi:10.1183/09031936.00065706 (2007).

4 Loh, P. R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet* **47**, 1385-1392, doi:10.1038/ng.3431 (2015).

5 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-375, S361-363, doi:10.1038/ng.2213
ng.2213 [pii] (2012).

6 Ferreira, M. A. Improving the power to detect risk variants for allergic disease by defining case-control status based on both asthma and hay fever. *Twin Res Hum Genet* **17**, 505-511, doi:10.1017/thg.2014.59 (2014).

7 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*, doi:10.1038/ng.3211 (2015).

8 Consortium, G. T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660, doi:10.1126/science.1262110 (2015).

9 Wells, A. *et al.* The anatomical distribution of genetic associations. *Nucleic Acids Res* **43**, 10804-10820, doi:10.1093/nar/gkv1262 (2015).

10 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).

11 Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-343, doi:10.1038/nature13835 (2015).

12 Fehrmann, R. S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* **47**, 115-125, doi:10.1038/ng.3173 (2015).

13 Thomsen, S. F., Kyvik, K. O. & Backer, V. Etiological relationships in atopy: a review of twin studies. *Twin Res Hum Genet* **11**, 112-120, doi:10.1375/twin.11.2.112 (2008).

14 Sanseau, P. *et al.* Use of genome-wide association studies for drug repositioning. *Nature biotechnology* **30**, 317-320, doi:10.1038/nbt.2151 (2012).

15 Bonder, M. J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* **49**, 131-138, doi:10.1038/ng.3721 (2017).

16 Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet* **9**,

436-447, doi:10.1161/CIRCGENETICS.116.001506 (2016).
17 Lev, S. *et al.* Identification of a novel family of targets of PYK2 related to Drosophila retinal degeneration B (rdgB) protein. *Mol Cell Biol* **19**, 2278-2288 (1999).
18 Odemuyiwa, S. O. *et al.* Cyclin-dependent kinase 5 regulates degranulation in human eosinophils. *Immunology* **144**, 641-648, doi:10.1111/imm.12416 (2015).
19 Kamen, L. A., Schlessinger, J. & Lowell, C. A. Pyk2 is required for neutrophil degranulation and host defense responses to bacterial infection. *J Immunol* **186**, 1656-1665, doi:10.4049/jimmunol.1002093 (2011).

ACKNOWLEDGMENTS

This research was conducted using the UK Biobank resource under Application Number 10074. Detailed acknowledgments are provided for each contributing study in the Supplementary Information.

AUTHOR CONTRIBUTIONS

Data collection and analysis in the contributing studies. AAGC study: M.A.F., M.C.M., S.C.D., L.M.B., P.J.T., N.G.M., D.L.D.; LifeLines study: J.M.V., G.H.K.; GENEVA study: H.B., E.R., M.H., A.F., N.N., H.S.,S.K., C.G., K.S., S.W.; GENUFAD studies: I.M., F.R., J.E-G., S.G., A.A., G.H., C.O.S., N.H., Y-A,L.; 23andMe study: C.T., D.A.H.; GERA study: J.D.H., J.S.W., R.B.M, E.J.; NTR study: Q.H., J-J,H., G.W., D.I.B.; CATSS, TWINGENE and SALTY studies: A.T., V.U., Y.L., P.K.E.M., C.A., R.K.; ALSPAC study: L.P.; HUNT study: B.M.B., L.F., M.E.G., J.B.N., W.Z., K.H., A.L., O.L.H., M.L., G.A., C.W.; UK Biobank study: L.P., M.A.F.

Methylation analysis: J.vD., D.I.B., R.J.

Biological and drug annotation: M.A.F., C.W.M., E.M., K.B., O.H., J.Z., J.A.R., J.B., B.B.

Quality control, meta-analysis, tables and figures: M.A.F.

Writing group: M.A.F., J.M.V., I.M., C.T., J.D.H., Q.H., A.T., V.U., J.vD., Y.L., J.E-G., B.M.B., J.B., S.C.D., S.W., P.K.E.M., R.J., E.J., Y-A.L., D.I.B., C.A., R.K., G.H.K., L.P.

Study design and management: M.A.F., D.A.H., B.M.B., S.W., P.K.E.M., R.J., E.J., Y-A.L., D.I.B., C.A., R.K., G.H. K., L.P.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology

ONLINE METHODS

List of risk variants reported to be associated with allergic disease in previous GWAS

We downloaded the full NHGRI-EBI GWAS catalog database [1] on January 19, 2017 (v1.0.1). We then identified SNP associations with a $P \leq 5x10^{-8}$ and that were reported for an allergic condition, specifically for which the "MAPPED_TRAIT" variable included the terms "allergic rhinitis", "allergic sensitization", "allergy", "asthma", "eczema" and "atopic march". Excluded associations were inspected to ensure that no relevant variants were missed by this filtering approach. After excluding two variants without a reference SNP (rs) ID (both in the MHC), there were 169 associations, including 144 unique rs IDs. We then used the --clump procedure in PLINK [2] and genotype data from individuals of European descent from the 1000 Genomes Project [3] (*n*=294, release 20130502_v5a) to reduce this list of 144 SNPs to variants in low linkage disequilibrium (LD) with each other (r^2<0.05), which are likely to represent statistically independent associations with allergic disease. After excluding five variants that were not polymorphic in Europeans (rs7212938, rs62176107, rs17218161, rs10056340, rs9273349), we identified 75 variants in low LD with each other. We then identified the earliest GWAS to report an association with each of these 75 variants (or with a SNP with r^2>0.05 with it) and used the year of publication to generate **Supplementary Fig. 2**.

Meta-analysis of allergic disease GWAS results conducted in 13 studies (*n*=360,838)

In each of 13 participating studies (**Supplementary Tables 1 and 2**), a GWAS was performed using an additive genetic model in individuals of European descent that reported suffering from asthma and/or hay fever and/or eczema (case-group, total *n*=180,129), against those who never reported suffering from any of these three conditions (control group, total *n*=180,709). A detailed description of the procedures used to identify cases and controls, as well as for SNP genotyping, imputation and association testing, is provided for each study in the **Supplementary Information**.

Prior to the meta-analysis, standard quality control (QC) filters were applied to results from individual studies (**Supplementary Table 1**). After QC, and restricting the analysis to SNPs present in at least the two largest studies (UK Biobank and 23andMe, combined *n*=256,623), results were available for 8,307,659 variants, of which most (89%) were available in >95% of the overall sample size. Intercept estimates from LD score regression analysis [4], which reflect inflation of test statistics that are likely due to technical biases, ranged between 1.00 and 1.16 (**Supplementary Table 1**). Results from individual studies were adjusted for the observed inflation by multiplying the square of the standard error of each genetic effect estimate by the respective LD score regression intercept. We then used METAL [5] to combine association results across studies using an inverse-variance-weighted, fixed-effects meta-analysis. *P*-values from the meta-analysis were further adjusted for the meta-analysis LD score regression intercept of 1.04. The genome-wide significance threshold was set at $3x10^{-8}$, as suggested previously for GWAS analyzing variants with MAF≥1% [6].

Identification of independent associations through approximate conditional analyses

For each chromosome, we identified all SNPs with a $P \leq 3x10^{-8}$, sorted these based on base pair position, and then grouped variants into the same locus if the distance between consecutive variants was <1Mb. Variants located >1 Mb from the previous genome-wide significant variant were assigned to a new locus. Next, for each of these loci, we identified statistically independent associations using approximate conditional analyses, as implemented in GCTA [7]. We refer to these as sentinel risk variants. In these analyses, LD calculations were based on a subset of 5,000 individuals from the UKBiobank study. Briefly, for each locus, we (1) identified the most significantly-associated SNP [*i*]; (2) adjusted the summary statistics of all SNPs in that locus by the effect of that top SNP; (3) identified the most significantly-associated SNP [*j*] that remained genome-wide significant in that locus; (4)

adjusted the summary statistics of all SNPs in that locus by the effects of SNPs *i* and *j*. We repeated this process until there were no SNPs associated with allergic disease at $P \leq 3x10^{-8}$ after adjusting for the effect of other, more strongly independently associated variants in that locus. Lastly, we estimated the LD between sentinel variants located in different risk loci (*i.e.* >1 Mb apart) and confirmed that the r^2 was always close to 0 (no pairs of sentinel variants with $r^2>0.02$).

Determining the novelty status of independent SNP associations with allergic disease

Previous GWAS identified 144 SNPs associated with the risk of various allergic conditions, which we grouped into 75 independent associations based on the LD between variants (see above). We used that information to classify each of our independent SNP associations into two major groups: located in known (<1Mb from any of those 144 previously reported associations; "KnownLocus") or new (>1Mb from those variants; "NewLocus") allergy risk loci. For the first group, we then estimated the LD between each sentinel variant identified in our study and all variant(s) reported in previous GWAS. If all reported variants had an $r^2<0.05$ with our sentinel variant, then our association was considered to represent a new risk variant in a known risk locus ("KnownLocus-NewVariant"). Alternatively, when at least one reported variant had an $r^2\geq0.05$, our association was considered to be a known risk variant in a known risk locus ("KnownLocus-KnownVariant"). The second major group of variants were located in new allergy risk loci. Within this group, we used the same approach just described to determine if our associations were novel when considering any disease or trait with genome-wide significant associations reported in the NHGRI-EBI GWAS catalog.

Comparison of risk allele frequencies between individuals suffering from a single allergic disease

By combining information from asthma, hay fever and eczema in the case-control definition used in

our GWAS, we expected our study design to improve power to identify risk variants shared between, but not specific to any of, the three diseases [8]. To understand if the associations discovered in our GWAS were indeed likely to represent risk factors shared across allergic diseases, we took advantage of the observation that not all affected individuals report allergic co-morbidities [9-11], and compared allele frequencies between three groups of adults: asthma-only cases (*n*=12,268), hay fever-only cases (*n*=33,305) and eczema-only cases (*n*=6,276). The studies that contributed to this analysis are indicated in **Supplementary Table 1** and described in detail in the **Supplementary Information**. We performed three sets of association analyses contrasting three non-overlapping groups of individuals: asthma-only (g1) vs. hay fever-only (g2); asthma-only (g1) vs. eczema-only (g3); and hay fever-only (g2) vs. eczema-only (g3). These analyses are statistically independent from the case-control analysis carried out as part of the GWAS, which facilitates interpretation of the results. For a given sentinel SNP, results from these analyses indicate if the risk allele is more (odds ratio [OR] >1) or less (OR<1) common in e.g. group 1 (g1) when compared to group 2 (g2). For example, if a SNP contributed similarly to the risks of asthma and hay fever but not eczema, then one would expect an OR~1 in the asthma-only vs. hay fever-only comparison, but an OR>1 in the asthma vs. eczema and hay fever vs. eczema analyses. The significance threshold for these analyses was set at 1.2×10^{-4}, which corresponds to a Bonferroni correction for the 136 SNPs and three sets of analyses performed (i.e. $P<0.05/(136 \times 3)$).

Association between sentinel risk variants and variation in allergy age-of-onset

There is considerable variation in the age allergic diseases are first reported, and this has been shown to be influenced by genetic risk factors [12]. We therefore studied the association between the sentinel variants identified in our GWAS and age-of-onset observed in the UK Biobank study (*n*=35,972). For each individual, we first considered the earliest age of any allergic disease (asthma or hay

fever/eczema; the latter two were covered by the same question, and so could not be differentiated) being reported. SNPs were tested for association with this phenotype, with sex and a SNP array variable included as covariates. The significance threshold used for this analysis was $3.6x10^{-4}$ (i.e. $P<0.05/136$). Because significant SNP associations with this broad age-of-onset phenotype could be driven by different risk allele frequencies amongst cases suffering from different individual conditions (for example, a FLG variant might be associated with earliest age-of-onset because it is more prevalent in eczema cases, which tends to precede the development of asthma and hay fever [13]), we repeated the analysis by considering individuals who had reported suffering only from a single disease: asthma-only (n=7,445), hay fever-only (n=4,232) and eczema-only (n=1,225). For a given SNP, differences in effect size (beta) between groups were quantified using the formula z = sigma / SE_sigma, where sigma = beta_groupA – beta_groupB, and SE_sigma = sqrt(SE_beta_groupA^2 + SE_beta_groupB^2), which follows a normal distribution.

Estimating the contribution of the sentinel variants to the heritability of asthma, hay fever and eczema

Five steps were involved. First, we performed a GWAS of the individual diseases in the HUNT study, which was not included in the discovery meta-analysis. The HUNT study is described in greater detail in the **Supplementary Information**. Briefly, based on self-reported questionnaire information, we identified 1,875 cases and 16,463 controls for the asthma GWAS; 6,939 cases and 12,844 controls for the hay fever GWAS; and 2,630 cases and 16,131 controls for the eczema GWAS. After quality control filters, we analyzed 7.6 million common variants (genotyped and imputed) for association with each individual phenotype. The genomic inflation factor (i.e. lambda) for these analyses were 1.049 for asthma, 1.078 for hay fever, and 1.041 for eczema. Second, for each of the three diseases, we

quantified the overall SNP-based heritabilities with LD score regression [4] using a subset of 1.2 million HapMap SNPs. To obtain a heritability estimate on the liability scale, we set the population prevalence to be the same as the sample prevalence, given that this was a population-based study. Third, we removed the 136 sentinel variants (and all correlated variants, r^2>0.05) from the individual disease GWAS results. Fourth, we re-estimated SNP-based heritabilities as described for step two, but now using the GWAS results without the 136 top associations. In the fifth and final step, the contribution of the 136 sentinel variants towards the heritability of each disease was calculated as the difference between the SNP-based heritability estimated in steps two (all SNPs) and four (without 136 top associations).

Identification of plausible target genes of sentinel risk variants

Two independent strategies were used to identify plausible target genes underlying the observed associations. By 'target gene' we mean a gene for which protein sequence and/or variation in transcription is associated with a sentinel risk variant or one of its proxies (r^2>0.8).

First, we used wANNOVAR [14] to identify genes containing non-synonymous SNPs amongst all variants in LD (r^2>0.8) with any sentinel risk variant. SNPs in LD with sentinel risk variants were identified using genotype data from individuals of European descent from the 1000 Genomes Project [3] (*n*=294, release 20130502_v5a).

Second, to identify genes with transcription levels associated with a sentinel risk variant or one of its proxies (r^2>0.8), we queried publicly available results from 39 published expression quantitative trait loci (eQTL) studies conducted in 19 tissues or cell types relevant to allergic disease (**Supplementary Table 12**). We used a conservative significance threshold to identify significant SNP-gene expression associations, specifically a $P<2.3\times10^{-9}$ for *cis* effects (<1 Mb). We selected this

threshold based on a Bonferroni correction that considers the total number of protein-coding genes (G) and the number of SNPs likely to have been tested per gene (M): $P<0.05/(GxM)$. G was set at 21,742, based on the GeneCards database[15], queried on October 19th, 2016. We approximate M to be 1,000, as indicated by others [16-18], and so the threshold becomes $P=0.05/(21{,}472 \text{ genes x } 1{,}000 \text{ SNPs per gene})=2.3x10^{-9}$. We did not use information from *trans* eQTLs to identify plausible target genes of sentinel risk variants, because often these are thought to involve indirect effects[19] (*e.g.* sentinel SNP influences the expression of a transcript in *cis*, which in turn affects the expression of many other genes in *trans*).

For each eQTL study, and within each study for each tissue, we created a list of SNPs associated with gene expression in cis at a $P<2.3x10^{-9}$. Then, for each gene in that study-tissue dataset, we used the --clump procedure in PLINK to reduced the list of expression-associated SNPs (which often included many correlated SNPs) to a set of 'sentinel eQTLs', defined as the SNPs with strongest association with gene expression and in low LD ($r^2<0.05$, LD window of 2 Mb) with each other. This procedure was repeated for each of the 94 study-tissue datasets listed in **Supplementary Table 12**. Finally, we identified as a likely target of a sentinel allergy risk variant any gene for which a sentinel eQTL in any of the 94 study-tissue datasets had an LD $r^2>0.8$ with the sentinel risk variant. That is, we only considered genes for which there was strong LD between a sentinel variant and a sentinel eQTL, which reduces the chance of spurious co-localization. We did not use statistical approaches developed to distinguish co-localization from shared genetic effects because these have very limited resolution at high LD levels ($r^2>0.8$) [20].

To help prioritize plausible target genes for functional validation in subsequent studies, we identified genes for which publicly available functional data supported not just the presence of chromatin interactions between an enhancer and a gene promoter (based on 5C[21], promoter capture Hi-

C[22], ChIA-PET[23] or *in situ* Hi-C[24] data), but also an association between variation in enhancer epigenetic marks and variation in gene transcription levels (based on PreSTIGE[25], H3K27ac enhancer and super-enhancer annotation [26], IM-PET[27] or FANTOM5[28] analyses). We considered data from immune cell types, lung and skin (**Supplementary Table 15**) and putative enhancers that overlapped a sentinel risk variant (or one of its strongly correlated proxies, r^2>0.95).

To identify genes that were unlikely to have been previously implicated in the pathophysiology of allergic disease, we performed the following PubMed query on May 19th, 2017: (asthma OR rhinitis OR eczema OR atopic OR dermatitis OR allergy OR allergi* OR hayfever OR "hay fever") AND (gene1 OR gene2 OR … OR gene244). The gene symbols approved by the HUGO Gene Nomenclature Committee (HGNC) for each of the target genes were inserted into the second part of that query. The search results were downloaded as an .xml file and the number of unique articles (based on PMID) listing a given gene symbol was counted using in-house scripts (results in **Supplementary Table 14**). To identify genes likely to have been implicated in immune-related processes, we repeated this approach but replaced the first part of the PubMed query with (immune OR immuni* OR immunol*).

Enrichment in tissue-specific gene expression

We used the TSEA approach [29] to identify tissues that were likely to be affected functionally by the biological effects of the sentinel risk variants. We implemented this approach locally using custom scripts. Specifically, for each of 25 broad tissue types studied by the GTEx consortium, we tested if genes with tissue-specific expression (based on a Specificity Index threshold [29] [pSI] of 0.05; listed in file TableS3_NAR_Dougherty_Tissue_gene_pSI_v3-1.txt, downloaded from http://genetics.wustl.edu/jdlab/psi_package/) were enriched amongst the list of plausible target genes, when compared to the rest of the genes in the genome. After excluding genes without a pSI value and

189 in the MHC, there were 112 plausible target genes and 17,671 background genes available for analysis.
190 To test if the plausible target genes were enriched for genes with specific expression in a given tissue,
191 we used Fisher's exact test (one-sided). To rule out the possibility that a significant enrichment could
192 arise because the list of plausible targets was enriched for genes with eQTLs, we repeated the analysis
193 after restricting the background gene list to a subset of 12,804 genes that were found to have eQTLs in
194 the same eQTL studies that were used to identify plausible target genes of sentinel variants.

195 We also tested if a significant enrichment in tissue-specific expression could be a general
196 feature of genes near sentinel risk variants, and not specific to the list of genes identified as plausible
197 targets. To address this possibility, we generated 1,000 arbitrary gene lists, each containing 112 random
198 genes instead of the plausible target genes. We selected genes at random from the 17,783 with an
199 available pSI value and not in the MHC, using three strategies. First, genes were randomly drawn from
200 allergy risk loci (+/- 1 Mb of a sentinel variant). To generate each list of random genes, for each non-
201 MHC allergy risk locus *L*, we randomly selected a locus *R* from the subset of non-MHC allergy risk
202 loci for which the number of genes available for selection was the same or greater than the actual
203 number of plausible target genes (*T)* selected for that locus *L*. Then, for that locus *R*, we selected *T*
204 genes at random from the available genes in that locus. This procedure was repeated for all non-MHC
205 allergy risk loci, ensuring that the same locus was not selected twice in a given random dataset.

206 In the second strategy, genes were randomly drawn from 2 Mb loci selected at random from the
207 genome. In this case, to generate each list of random genes, we first partitioned the autosomes
208 (excluding the MHC) into 1,430 consecutive 2 Mb loci, and counted how many genes with an available
209 pSI value were present in each of these loci. Then, for each non-MHC allergy risk locus *L*, we
210 randomly selected a locus *R* from the subset of 2 Mb loci for which the number of genes available for
211 selection satisfied the following criteria: (1) was the same or greater than the actual number of plausible

target genes (*T)* selected for that locus *L*; and (2) matched (within 10%) the number of genes available for selection for that locus *L*. This was important to ensure that the randomly selected locus *R* was comparable to the allergy risk locus *L* in terms of the number of genes available for selection. Then, for that locus *R*, we selected *T* genes at random from the available genes in that locus.

In the third and final strategy, we simply selected genes at random from all 17,783 non-MHC genes with an available pSI value, ignoring where the genes were located in the genome. As a result, for a given random list, the genes selected could only be in close proximity to other genes in that same list by chance alone.

The same approach used to test the enrichment in tissue-specific expression for the plausible target genes was then used to analyze each of the 1,000 lists of random genes. For each of these lists, the smallest *P*-value observed across all 25 tissues tested was retained (P_{min}). The proportion of random gene lists (out of 1,000) with a P_{min} that was the same or lower than the enrichment *P*-value observed with the plausible target genes (P_{obs}) was then calculated. This corresponds to the probability of exceeding that enrichment when analyzing the random gene lists, after correcting for the 25 tissues tested. As we did for the analysis of the plausible target genes, we repeated the generation and analysis of random gene lists after restricting the genes available for selection (and the background gene list) to the subset of genes with a known eQTL.

Enrichment in tissue-specific SNP heritability

Finucane et al. [30] developed an approach to identify tissues likely affected by the functional effects of disease risk variants, called stratified LD score regression. This approach quantifies the contribution of SNPs located in tissue-specific regulatory annotations to the overall disease heritability. As such, it does not require the identification of likely target genes of allergy risk variant and considers all SNPs in

the genome, not just those with a genome-wide significant association with disease risk. Specifically, up to four histone marks (H3K4e1, H3K4me3, H3K9ac and H3K27ac) measured by the ENCODE project are used to define regulatory annotations (*e.g.* enhancers) in 100 different cell types. SNPs that overlap these regulatory annotations are then identified and their contribution as a group to the disease heritability quantified. As recommended by Finucane et al. [30], we ranked cell types based on the *P*-value of the regression coefficient, rather than the *P*-value of total enrichment. To ensure that significant SNP heritability enrichments were not explained by the effects of sentinel variants, we removed the top SNPs (and any variants with r^2>0.05 with these) from the meta-analysis GWAS results and repeated the LD score regression analysis.

Enrichment of biological processes

To identify biological processes enriched amongst the non-MHC target genes, we used GeneNetwork [31]. With this approach, gene sets originally included in a given GO biological process (BP) were expanded to include other genes based on a 'guilt-by-association' procedure [31]. After excluding BPs with <10 or >500 genes, 3,770 BPs were available for analysis. For each BP, we tested its enrichment amongst the list of plausible target genes as follows. First, we downloaded a gene set file containing *z*-scores for each of 19,976 unique genes in the genome from http://129.125.135.180:8080/GeneNetwork/resources/ontology?ontology=GO_BP&term=[pathway], where 'pathway' was replaced with the actual name of the BP being tested (*e.g.* "GO:0000002"). The z-score for gene X in that file reflects the probability that gene X is part of that BP. Second, we compared the distribution of z-scores between the list of plausible target genes (107 non-MHC genes were in the GeneNetwork gene set files, and so were available for analysis) and a background gene list of 18,193 genes (obtained after excluding MHC genes, the 107 plausible target genes and genes not listed in

GENCODE release 19), using a one-sided Wilcoxon rank-sum test. The *P*-value from this test represents the probability that genes in that BP are enriched amongst the list of plausible target genes, when compared to the background gene list.

As for the enrichment analysis of tissue-specific expression, we estimated how often a BP enrichment observed with the list of plausible target genes would be expected had we sampled genes at random from the allergy risk loci or from random loci. This analysis addresses the possibility that an observed enrichment might not be a specific feature of the plausible target genes identified but instead a general feature of genes located near sentinel allergy risk variants, or simply in close proximity to each other. We used the same three strategies described above to generate 1,000 random gene lists, sampling from the 18,300 non-MHC with an available z-score and in GENCODE release 19. To determine if using eQTL information to identify plausible target genes could have biased the enrichment analysis, we generated and analysed random gene lists after restricting the genes available for selection to the subset with known eQTLs (12,913), but found very similar results (not shown).

Common traits and diseases associated with allergic disease risk variants

We first identified all variants in LD (r^2>0.8) with a sentinel risk variant using data from Europeans of the 1000 Genomes Project [3] (*n*=294, release 20130502_v5a), and extracted any associations with these reported in the NHGRI-EBI GWAS catalog database [1] (queried on December 13, 2016) or by Astle et al. [32], a large GWAS of blood cell counts (*n*=173,480). To complement this analysis, we estimated the SNP-based genetic correlation between our GWAS and results reported for 229 common traits or diseases, using LD Hub [33]. In these analyses, results from our meta-analysis were not corrected for the LD score intercept, either at the study level or after the meta-analysis.

Identification of target genes with drugs considered as drug targets for human diseases

To identify genes that encode transcripts that are targets of drugs considered for clinical development, we queried the Thomson Reuters Cortellis™ Drug database between November 7 and 15, 2016, which included 63,417 drugs. The drug search was carried out individually for each gene. First, a search query was built based on the following format: HGNC approved gene name OR alias_1 OR … OR alias_N. Gene name aliases were obtained from the Bioconductor annotation package org.Hs.eg.db. For example, to find drugs that target *IL6R*, the search query used was: "CD126" OR "IL-6R-1" OR "IL-6RA" OR "IL6Q" OR "IL6RA" OR "IL6RQ" OR "gp80" OR "IL6R" OR "interleukin 6 receptor". Second, after running the search query, results were filtered based on the ascribed "Target-based Actions", keeping only entries that corresponded to the gene name or an alias. For example, of the 65 results obtained with the *IL6R* query above, only for 20 did the target-based action mention *IL6R* or an alias. Third, drug results were downloaded, and the gene and respective drug allocated to one of three groups: (1) gene with at least one drug considered for the treatment of allergic diseases (15 genes); (2) gene considered for the treatment of immune-related conditions, but not allergic diseases specifically (8 genes); and (3) gene considered for the treatment of other conditions (26 genes).

Directional effect of the allergy protective allele on target gene expression

In an attempt to predict if existing drugs would be expected to attenuate or exacerbate allergic symptoms, we compared the effect on gene expression between the allergy protective allele and the existing drug. We acknowledge that this is a simplistic comparison, because it assumes that the directional effect of the protective allele is not tissue- or context-dependent, which is true for most but not all expression-associated SNPs [34-36].

To determine if the allergy protective allele of a sentinel variant was associated with higher or

lower target gene expression, we focused on the subset of target genes identified via an eQTL (see above). This was straightforward to assess when the sentinel SNP and the expression-associated SNP were the same variant: for example, if the allergy-protective allele had a negative effect (e.g. beta or z-score) on gene expression in the published eQTL study, then that allele was associated with lower gene expression. On the other hand, when the two SNPs did not correspond to the same variant, but were in high LD (r^2>0.8) with each other, we first determined which allele of the expression-associated SNP was on the same haplotype as the allergy-risk allele. Then we used that allele to infer the direction of effect of the allergy-risk allele on gene expression.

Modulation of target gene methylation by environmental risk factors

We first tested if variation in DNA CpG methylation was associated with variation in target gene expression, independently of SNP effects, using data from the Biobank-based Integrative Omics Study (BIOS) consortium that is described in detail elsewhere [37,38]. Methylation and expression levels in whole-blood samples (*n*=2,101) were quantified respectively with Illumina Infinium HumanMethylation450 BeadChip Kit arrays and RNA-seq (2x50bp paired-end, Hiseq2000, >15M read pairs per sample). For each target gene, we identified CpG sites in *cis* (<250 Kb from gene) for which methylation levels were significantly associated with gene expression levels (FDR<5%), after adjusting the methylation levels for methyation-associated SNPs and expression levels for expression-associated SNPs. Such CpG sites, called *cis*-eQTMs, were identified in a previous study [37] and downloaded from http://genenetwork.nl/biosqtlbrowser. For most genes, there were multiple *cis*-eQTMs, and so we selected the CpG site most strongly associated with variation in gene expression for downstream analyses.

Next, we tested the association between methylation levels at these sentinel CpGs with five

established risk factors for allergic disease using data from unrelated individuals of the Netherlands Twin Register (NTR) study, which was included in the BIOS consortium studies [37,38]. The risk factors tested were current smoking (*n*=1,221), maternal smoking (*n*=637), BMI (*n*=1,214), birth weight (*n*=1,015) and number of older siblings (*n*=775). Information on BMI and current smoking was collected as part of the NTR biobank project [39] at blood draw. Birth weight was obtained in multiple NTR surveys as previously described [40]. Maternal smoking during pregnancy was measured in NTR Survey 10 (data collection in 2013) with the following question: Did your mother ever smoke during pregnancy? with answer categories: no, yes, I don't know. Information on the number of older siblings was obtained through self-report in NTR surveys 2, 3 and 6. For twin pairs, the answers were checked for consistency and missing data for one twin were supplemented with data from the co-twin where possible. Linear or logistic regression was used to test the association between methylation (β-value) and individual risk factors, with the following variables included as covariates: sex, age at blood sampling, methylation array row, bisulphite plate and white blood cell percentages (% neutrophils, % monocytes, and % eosinophils). The association with maternal smoking was tested while also adjusting for smoking status.

Data availability

Summary statistics of the meta-analysis without the 23andMe study will be made publicly available at the time of publication. The full GWAS summary statistics for the 23andMe discovery data set will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please contact David Hinds (dhinds@23andme.com) for more information and to apply to access the data.

References

1 Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001-1006, doi:10.1093/nar/gkt1229 (2014).

2 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575, doi:S0002-9297(07)61352-4 [pii] 10.1086/519795 (2007).

3 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

4 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*, doi:10.1038/ng.3211 (2015).

5 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:btq340 [pii] 10.1093/bioinformatics/btq340 (2010).

6 Fadista, J., Manning, A. K., Florez, J. C. & Groop, L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *Eur J Hum Genet* **24**, 1202-1205, doi:10.1038/ejhg.2015.269 (2016).

7 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-375, S361-363, doi:10.1038/ng.2213 ng.2213 [pii] (2012).

8 Ferreira, M. A. Improving the power to detect risk variants for allergic disease by defining case-control status based on both asthma and hay fever. *Twin Res Hum Genet* **17**, 505-511, doi:10.1017/thg.2014.59 (2014).

9 Pinart, M. *et al.* Comorbidity of eczema, rhinitis, and asthma in IgE-sensitised and non-IgE-sensitised children in MeDALL: a population-based cohort study. *The Lancet. Respiratory medicine* **2**, 131-140, doi:10.1016/S2213-2600(13)70277-7 (2014).

10 Gough, H. *et al.* Allergic multimorbidity of asthma, rhinitis and eczema over 20 years in the German birth cohort MAS. *Pediatr Allergy Immunol* **26**, 431-437, doi:10.1111/pai.12410 (2015).

11 Mortz, C. G., Andersen, K. E., Dellgren, C., Barington, T. & Bindslev-Jensen, C. Atopic dermatitis from adolescence to adulthood in the TOACS cohort: prevalence, persistence and comorbidities. *Allergy* **70**, 836-845, doi:10.1111/all.12619 (2015).

12 Sarnowski, C. *et al.* Identification of a new locus at 16q12 associated with time to asthma onset. *J Allergy Clin Immunol* **138**, 1071-1080, doi:10.1016/j.jaci.2016.03.018 (2016).

13 Dharmage, S. C. *et al.* Atopic dermatitis and the atopic march revisited. *Allergy* **69**, 17-27, doi:10.1111/all.12268 (2014).

14 Chang, X. & Wang, K. wANNOVAR: annotating genetic variants for personal genomes via the web. *J Med Genet* **49**, 433-436, doi:10.1136/jmedgenet-2012-100918 (2012).

15 Rebhan, M., Chalifa-Caspi, V., Prilusky, J. & Lancet, D. GeneCards: integrating information about genes, proteins and diseases. *Trends in genetics : TIG* **13**, 163 (1997).

16 Davis, J. R. *et al.* An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *Am J Hum Genet* **98**, 216-224, doi:10.1016/j.ajhg.2015.11.021 (2016).
17 Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773-777, doi:10.1038/nature08903 (2010).
18 Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506-511, doi:10.1038/nature12531 (2013).
19 Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-1243, doi:10.1038/ng.2756 (2013).
20 Chun, S. *et al.* Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat Genet* **49**, 600-605, doi:10.1038/ng.3795 (2017).
21 Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109-113, doi:10.1038/nature11279 (2012).
22 Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet* **47**, 598-606, doi:10.1038/ng.3286 (2015).
23 Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98, doi:10.1016/j.cell.2011.12.014 (2012).
24 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).
25 Corradin, O. *et al.* Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res* **24**, 1-13, doi:10.1101/gr.164079.113 (2014).
26 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).
27 He, B., Chen, C., Teng, L. & Tan, K. Global view of enhancer-promoter interactome in human cells. *Proc Natl Acad Sci U S A* **111**, E2191-2199, doi:10.1073/pnas.1320308111 (2014).
28 Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461, doi:10.1038/nature12787 (2014).
29 Wells, A. *et al.* The anatomical distribution of genetic associations. *Nucleic Acids Res* **43**, 10804-10820, doi:10.1093/nar/gkv1262 (2015).
30 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).
31 Fehrmann, R. S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* **47**, 115-125, doi:10.1038/ng.3173 (2015).
32 Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429 e1419, doi:10.1016/j.cell.2016.10.042 (2016).
33 Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and

genetic correlation analysis. *Bioinformatics* **33**, 272-279, doi:10.1093/bioinformatics/btw613 (2017).

34 Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* **44**, 502-510, doi:10.1038/ng.2205 (2012).

35 Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949, doi:10.1126/science.1246949 (2014).

36 Fu, J. *et al.* Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS Genet* **8**, e1002431, doi:10.1371/journal.pgen.1002431 (2012).

37 Bonder, M. J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* **49**, 131-138, doi:10.1038/ng.3721 (2017).

38 Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nat Genet* **49**, 139-145, doi:10.1038/ng.3737 (2017).

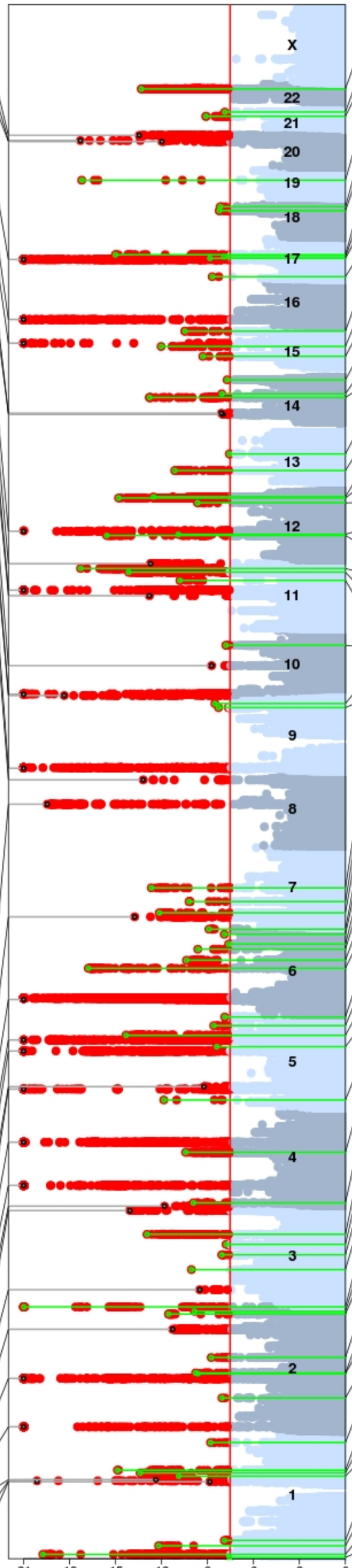
39 Willemsen, G. *et al.* The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. *Twin Res Hum Genet* **13**, 231-245, doi:10.1375/twin.13.3.231 (2010).

40 Tsai, P. C. *et al.* DNA Methylation Changes in the IGF1R Gene in Birth Weight Discordant Adult Monozygotic Twins. *Twin Res Hum Genet* **18**, 635-646, doi:10.1017/thg.2015.76 (2015).

35 known loci

Target genes	N target genes	N sentinel variants
STMN3	1	1
[ZNF217]	0	1
[NFATC2]	0	1
ORMDL3,CCR7,RP11-94L15.2,ZPBP2,GSDMB	7	4
CLEC16A-[]-RMI2,RMI2-[]--LITAF	0	2
AAGAB	1	1
FOXA1-[]--TTC6	0	1
NFKBIA,PPP2R3C,KIAA0391,FAM177A1	4	1
NAB2,STAT6,SUOX,RPS26,PA2G4,MYL6B,ERBB3	8	2
KIRREL3-AS3---[]--ETS1	0	1
WNT11--[]-LRRC32	0	3
OVOL1,SNX32,RP11-770G2.5,EFEMP2	4	1
[ZNF365]	0	1
GATA3---[]---SFTA1P,GATA3--[]SFTA1P	0	6
IL2RA	1	2
ERMP1	1	3
[MYC]	0	1
MIR5708--[]--ZBTB10	0	1
ITGB8--[]-ABCB5,[ITGB8]	0	2
HLA-C,MCCD1P1,HLA-DQA1,GPANK1,PRRC2A	5	7
SLC22A4,SLC22A5,C5orf56,IL13	4	3
CAMK4,SLC25A46,TSLP,CTC-551A13.2,WDR36	5	4
PTGER4	1	1
IL7R	1	1
KIAA1109	1	2
TLR1	1	1
FBXO45-[]-CEP19	0	1
RP11-132N15.2,RP11-132N15.1,BCL6	3	4
[GLB1]	0	1
BOLL,RFTN2,MARS2,PLCL1	4	1
MFSD9,IL18R1,IL1RL1,IL18RAP	4	2
[LINC00299]	0	1
IL6R	1	1
THEM4,FLG	2	3
RPRD2,C1orf54,TARS2	3	1

Figure legend

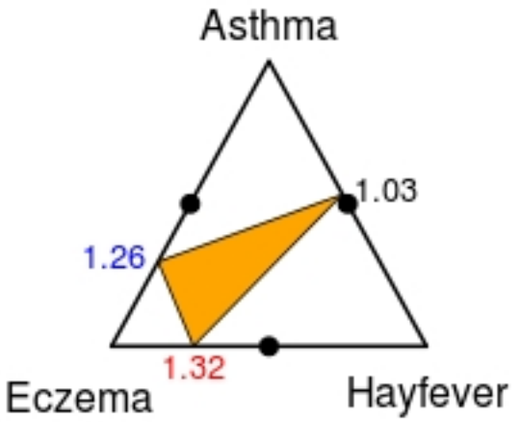


64 new loci

N sentinel variants	N target genes	Target genes
1	11	*C22orf46,PMM1,NHP2L1,CCDC134,PHF5A,CSDC2,MEI1*
1	0	[SIK1]
1	0	[RUNX1]
1	0	SLC7A10-[]-CEBPA
1	1	*PIGN*
1	1	*DYNAP*
1	1	*GNGT2*
1	2	*MAP3K14-AS1,SPATA32*
1	0	[STAT5B]
1	1	*ALOX15*
1	2	*IQGAP1,CRTC3*
1	1	*RP11-554D20.1*
1	2	*RTF1,ITPKA*
1	0	RCOR1-[]-TRAF3
1	0	JDP2-[]-BATF
1	0	[RAD51B]
1	0	PIBF1-[]-KLF5
1	0	[FOXO1]
1	10	*SBNO1,ABCB9,ARL6IP4,OGFOD2,PITPNM2,RP11-282O18.6*
1	1	*SPPL3*
1	1	*SH2B3*
1	1	*AQP5*
1	1	*HDAC7*
1	0	DDX6-[]-CXCR5
1	2	*PPP2R1B,SIK2*
1	0	SESN3--[]-FAM76B
1	3	*ARL3,TMEM180,ACTR1A*
1	0	C9orf114-[]-LRRC8A
1	1	*FBXW2*
1	1	*GSAP*
1	0	C7orf72-[]-IKZF1
1	1	*JAZF1*
1	0	RNASET2-[]-MIR3939
1	0	[ARID1B]
1	1	*TNFAIP3*
1	0	[PTPRK]
1	0	[ATG5]
1	0	[BACH2]
1	2	*RGS14,RAB24*
1	0	MIR3142-[]-MIR146A
2	2	*NDFIP1,HDAC3*
1	0	[TNFAIP8]
1	1	*FAM105A*
1	0	[MANBA]
1	0	STX18--[]-MSX1
1	1	*RASA2*
1	4	*SLC15A2,EAF2,IQCB1,CD86*
1	1	*SENP7*
1	0	LINC00870--[]-RYBP
1	0	[D2HGDH]
1	1	*INPP5D*
1	0	CCL20-[]-DAW1
1	1	*ARHGAP15*
1	0	[IL1B]
2	0	BCL2L11--[]--ANAPC1
1	0	[LOC339807]
1	0	[ITPKB]
2	0	TNFSF18--[]-TNFSF4
1	1	*CD247*
1	7	*FCER1G,B4GALT3,ADAMTS4,PPOX,F11R,USF1,TOMM40L*
1	0	SFPQ-[]-ZMYM4
1	1	*RUNX3*
1	2	*RERE,RP5-1115A15.1*
1	1	*TNFRSF14*

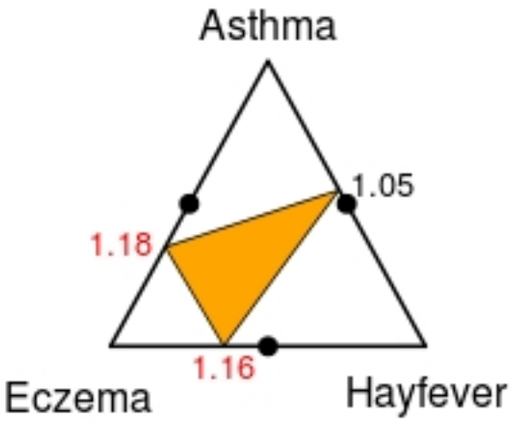
rs61816761:A (chr1:152 Mb)

[FLG]



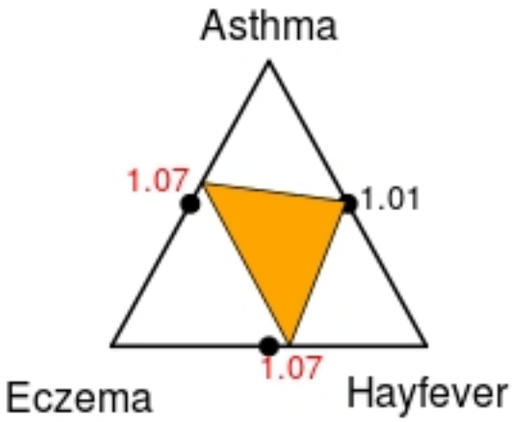
rs12123821:T (chr1:152 Mb)

RPTN-[]-HRNR



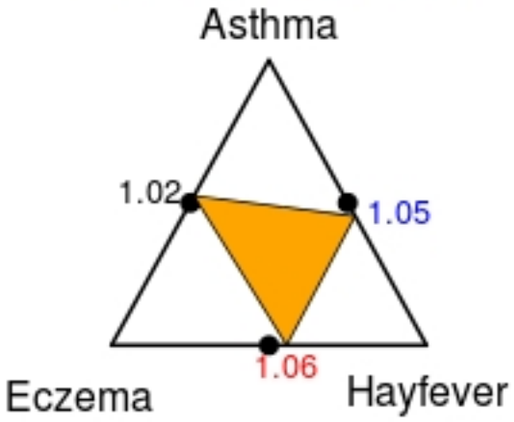
rs12470864:A (chr2:103 Mb)

IL1RL2-[]-IL18R1



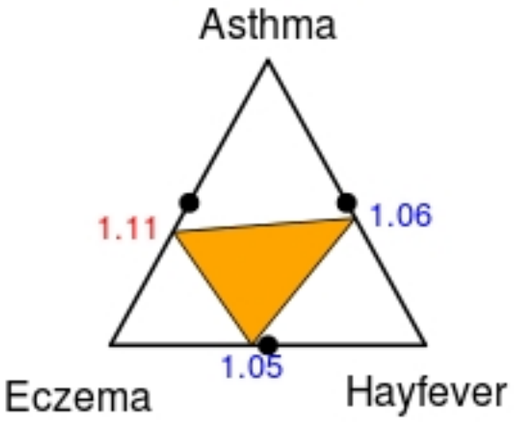
rs6594499:C (chr5:110 Mb)

WDR36-[]-CAMK4



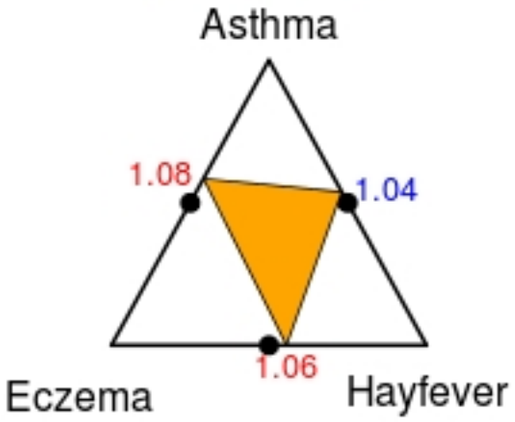
rs61839660:T (chr10:6 Mb)

[IL2RA]



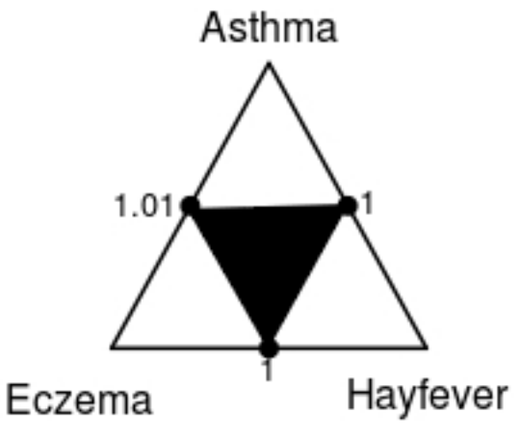
rs921650:A (chr17:38 Mb)

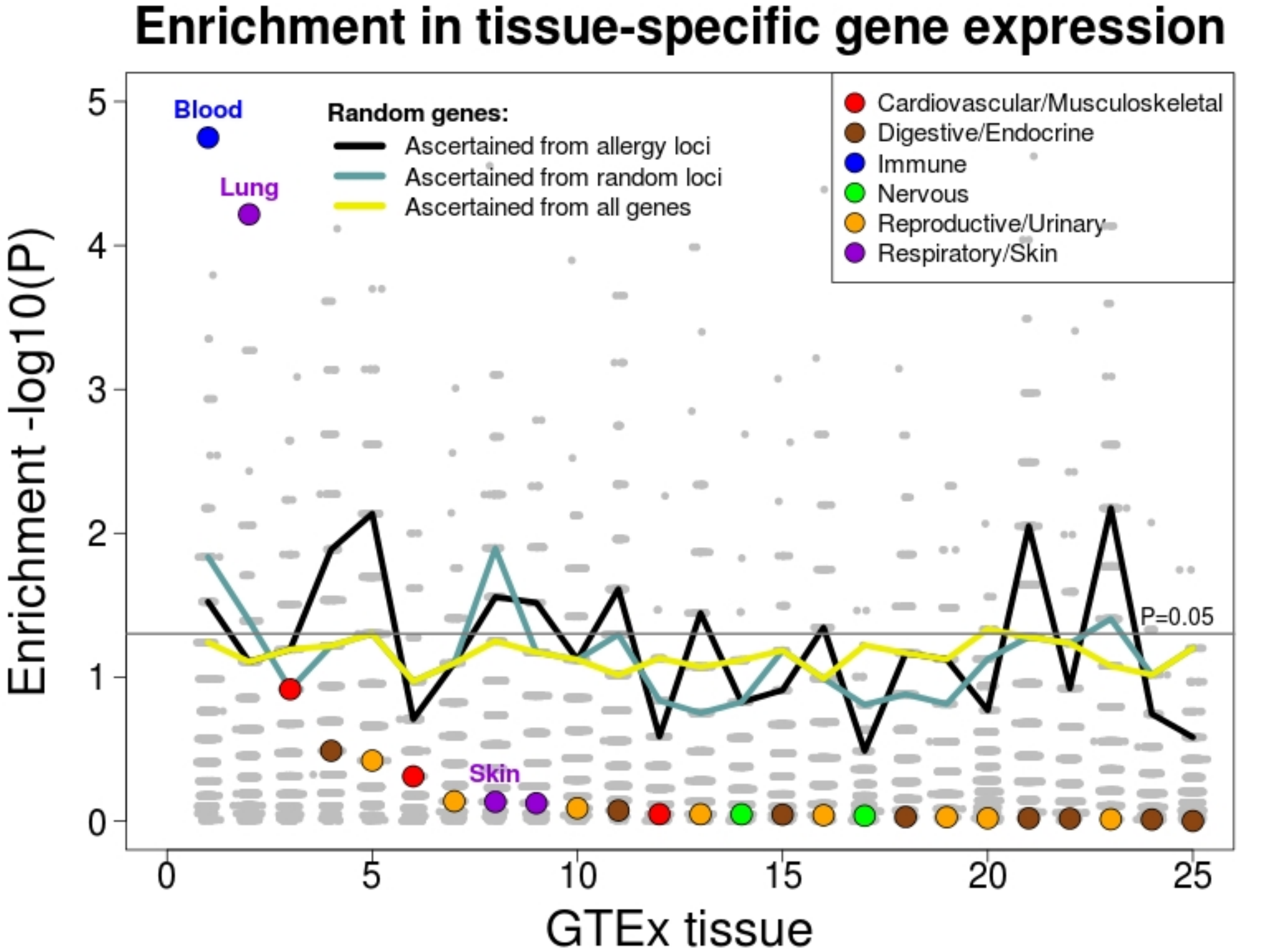
[GSDMB]

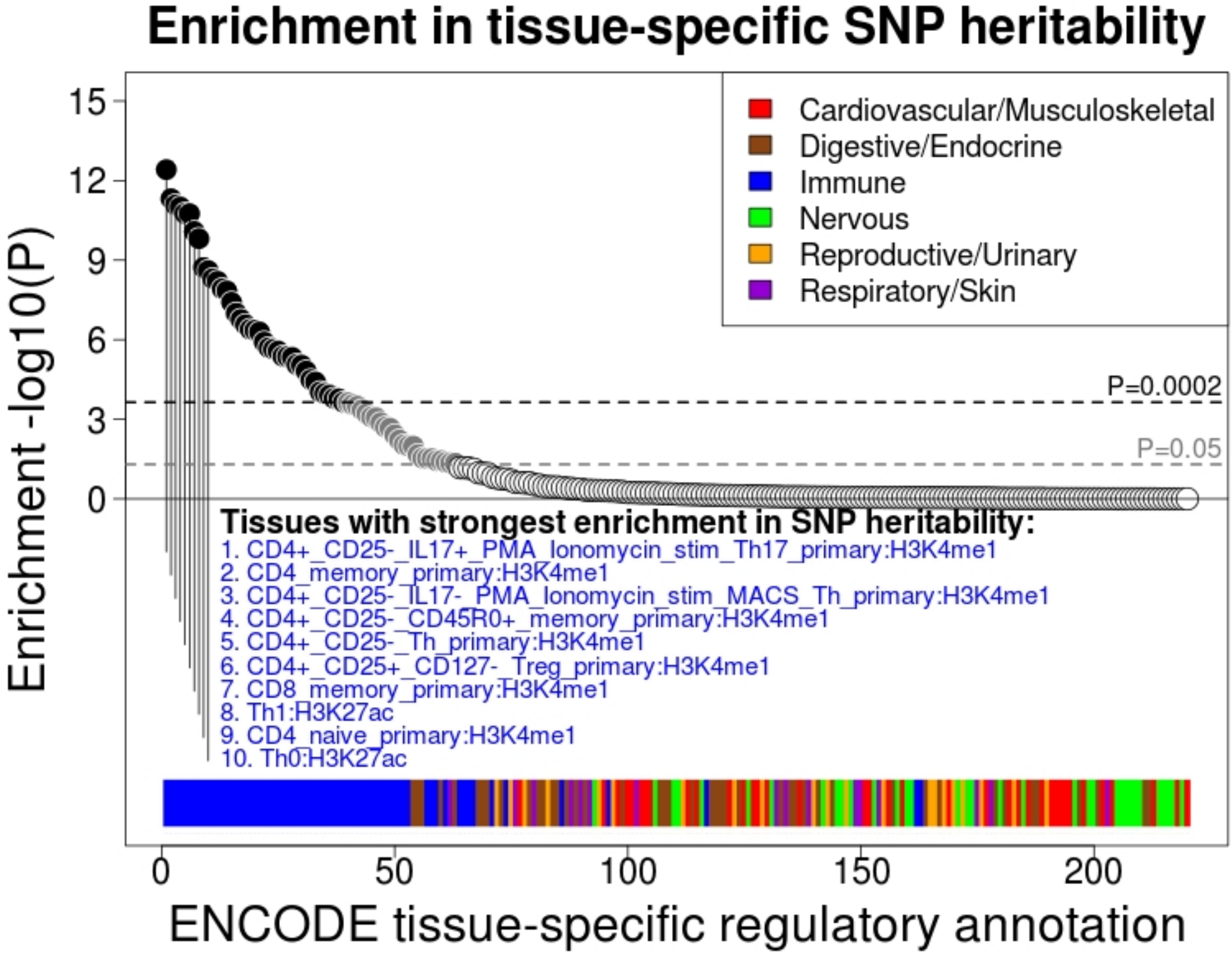


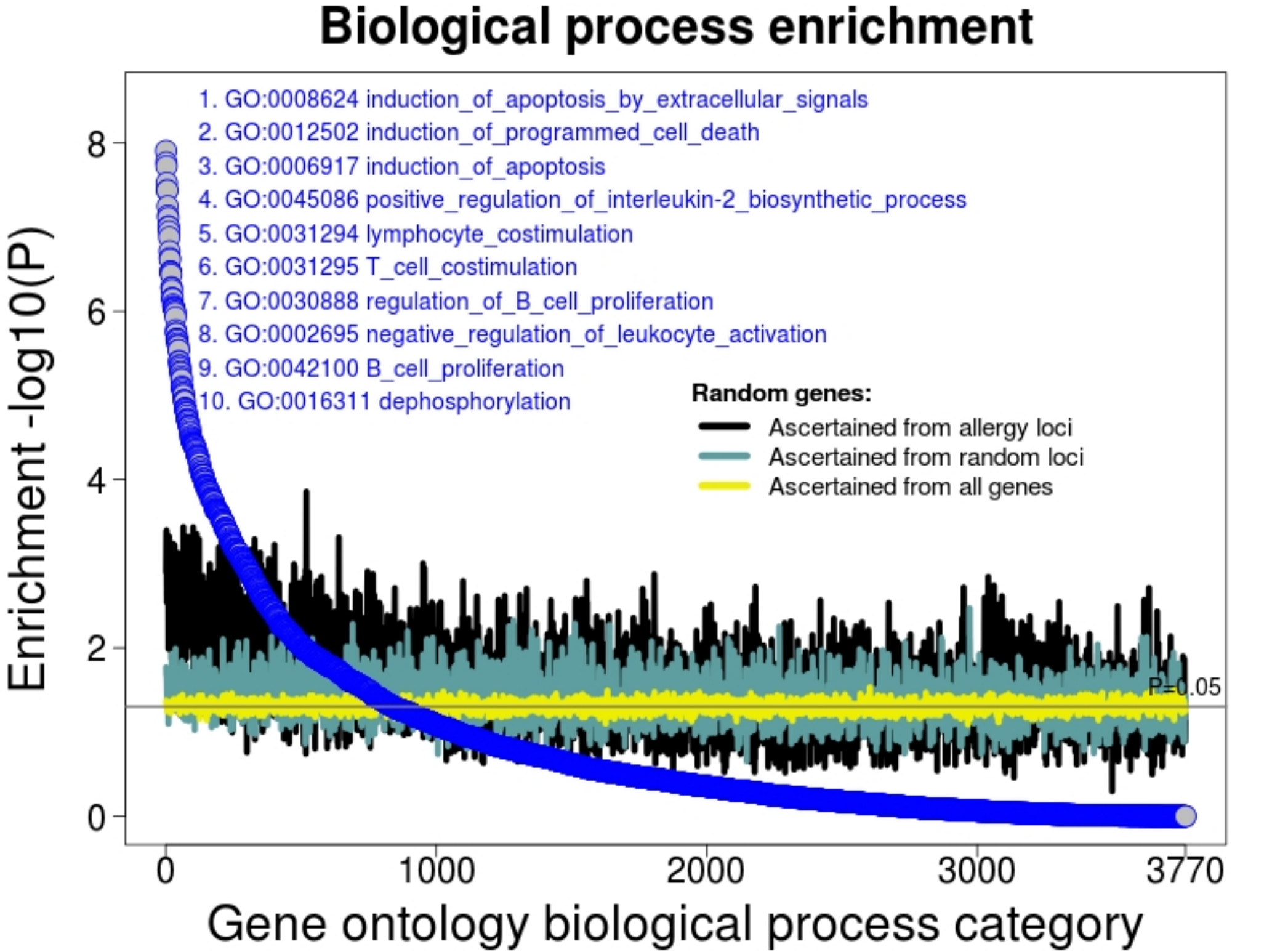
rs2228145:C (chr1:154 Mb)

[IL6R]




Enrichment in tissue-specific gene expression
Blood
Lung
Skin
Random genes:
Ascertained from allergy loci
Ascertained from random loci
Ascertained from all genes
Cardiovascular/Musculoskeletal
Digestive/Endocrine
Immune
Nervous
Reproductive/Urinary
Respiratory/Skin
P=0.05
Enrichment -log10(P)
GTEx tissue
0
1
2
3
4
5
0
5
10
15
20
25


Enrichment in tissue-specific SNP heritability
Cardiovascular/Musculoskeletal
Digestive/Endocrine
Immune
Nervous
Reproductive/Urinary
Respiratory/Skin
P=0.0002
P=0.05
Enrichment -log10(P)
15
12
9
6
3
0
Tissues with strongest enrichment in SNP heritability:
1. CD4+_CD25-_IL17+_PMA_Ionomycin_stim_Th17_primary:H3K4me1
2. CD4_memory_primary:H3K4me1
3. CD4+_CD25-_IL17-_PMA_Ionomycin_stim_MACS_Th_primary:H3K4me1
4. CD4+_CD25-_CD45R0+_memory_primary:H3K4me1
5. CD4+_CD25-_Th_primary:H3K4me1
6. CD4+_CD25+_CD127-_Treg_primary:H3K4me1
7. CD8_memory_primary:H3K4me1
8. Th1:H3K27ac
9. CD4_naive_primary:H3K4me1
10. Th0:H3K27ac
0
50
100
150
200
ENCODE tissue-specific regulatory annotation


Biological process enrichment
1. GO:0008624 induction_of_apoptosis_by_extracellular_signals
2. GO:0012502 induction_of_programmed_cell_death
3. GO:0006917 induction_of_apoptosis
4. GO:0045086 positive_regulation_of_interleukin-2_biosynthetic_process
5. GO:0031294 lymphocyte_costimulation
6. GO:0031295 T_cell_costimulation
7. GO:0030888 regulation_of_B_cell_proliferation
8. GO:0002695 negative_regulation_of_leukocyte_activation
9. GO:0042100 B_cell_proliferation
10. GO:0016311 dephosphorylation
Random genes:
Ascertained from allergy loci
Ascertained from random loci
Ascertained from all genes
P=0.05
Enrichment -log10(P)
0
2
4
6
8
0
1000
2000
3000
3770
Gene ontology biological process category